1997

# Identification and characterization of proteins binding to regulatory elements of the Hoxa-5 spatial-specific enhancer

Tamara Mann Nowling
*Iowa State University*

# INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**UMI**

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

# Identification and characterization of proteins binding to regulatory elements of the *Hoxa-5* spatial-specific enhancer

by

Tamara Mann Nowling

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Molecular, Cellular and Developmental Biology

Major Professors: Christopher Tuggle and Marit Nilsen-Hamilton

Iowa State University

Ames, Iowa

1997

Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Tamara Mann Nowling

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Co-major Professor

Signature was redacted for privacy.

Co-major Professor

Signature was redacted for privacy.

For the Major Program

Signature was redacted for privacy.

For the Graduate college

For my Parents for their continued love and support. For my husband, Mark, for always being there for me and for his unconditional love, support and sense of humor and for my son, Duncan, whose smiles and giggles never fail to brighten my day, I hope one day my research will help to make his world a better place. And a special dedication to my Grandfather, Wally, who I know is watching over me.

TABLE OF CONTENTS

# GENERAL INTRODUCTION

## Dissertation Organization

My dissertation is centered around the investigation into how the gene *Hoxa-5* expression is regulated transcriptionally during development. Therefore, I will begin by giving a basic introduction into what is known about transcriptional regulation of gene expression. I will relate this information to genes expressed during embryogenesis, specifically the Homeobox (Hox) genes. Following this introduction I will give a literature review which will include a comparison of embryo patterning in two model systems, *Drosophila* and *Xenopus*, and some of the genes involved. One set of these genes includes the Hox genes which will be discussed next followed by an in depth literature review of *Hoxa-5* (distinction between a Hox gene and a Hox protein will be made as follows; genes will be capitalized and italicized, proteins will be in lower case and not italicized.) The first paper contains results obtained by Dr. Wen Zhou, Dr. Chi Nguyen-Huu and Dr. Chris Tuggle, that identify specific cis regulatory elements (Figures 1 and 2) and my results showing specific binding of trans-acting factors to one of the cis-acting elements, the brachial spinal cord (BSC) element. In the second paper I characterize the specific binding to the BSC element and identify specific binding of trans-acting factors to a second cis regulatory element, the upper cervical repressor (UCR) element. I will then give a general conclusion of the results presented in the papers and discuss what these results mean, as well as discuss possible ways to further characterize these binding factors and to identify these factors. In addition, I have included an appendix describing experiments that identify a cDNA encoding a new mouse protein that is a putative regulator of *Hoxa-5*.

## Introduction

The regulation of gene expression is a widely-studied area. Many aspects are understood in a general sense, but understanding the specific regulation of individual genes is important. What is understood are the players involved which simply include cis-acting regulatory elements and the factors that bind these elements, referred to as transcription factors. The cis-acting elements and trans-acting factors interact in order to regulate the promoter of a gene. Cis-acting elements can be sequences within the promoter, such as the TATA box consensus sequences which bind general transcription factors, proximal sequences near the promoter, or distal sequences which can be located at great distances from the promoter. The proximal and distal elements are generally enhancer elements defined by their ability to regulate the expression of a gene in an orientation-independent manner. These enhancer elements usually bind specific transcription factors. The general transcription factors mentioned earlier

are involved in basal transcription of genes from all promoters and include polymerase, as well as the transcription factors that bind the TATA box. The factors that bind the proximal and distal elements are often specific for certain promoters. These factors can be expressed in most cells and may be involved in the regulation of a wide array of genes or have a spatial-specific pattern of expression and act in specific or multiple cell types (reviewed in Biggin and Tjian, 1989). These spatial-specific factors include the homeodomain (Hox) proteins which are involved in controlling embryonic development.

Besides the homeodomain proteins there are several other classes of transcription factors including the helix-turn-helix proteins, the zinc finger proteins, the steroid receptors, the leucine zipper proteins, the helix-loop-helix proteins, and the β-sheet proteins. The proteins in these families are grouped according to how they bind DNA, which usually involves an α-helix but could be by another secondary structure. Because the members of these families bind using the same binding motif, they also bind similar sequences. Side chains off of the secondary structures and residues outside the secondary structures account for some recognition of different sites (reviewed in Pabo and Sauer, 1992). But how are different members of a family that recognize similar sequences involved in regulating transcription from a specific promoter? One of the biggest factors is how the different cis elements involved in regulating a promoter are arranged. In addition, many proteins bind to DNA as dimers, homodimers or heterodimers depending on what other proteins are present in that cell. Therefore, a protein may dimerize with itself or with other proteins, what a protein dimerizes with is dependent upon what other proteins are present which is ultimately dependent upon the cell type. Each dimer is selective, not only for the sequence it binds, but also for proteins binding to neighboring cis elements (Jones, 1990; Lamb and McKnight, 1991; Mitchell and Tjian, 1989).

As mentioned previously, the Hox genes are expressed during embryonic development in a spatially-specific manner. There are a number of Hox genes in vertebrates and their spatial expression patterns overlap. The overlapping expression of homeodomain proteins results in a combination of different homeodomain proteins in different cell types. It is this combination of proteins that determines the positive or negative regulation of genes these proteins regulate (reviewed in Biggin and Tjian, 1989). This is true in *Drosophila* and seems to be the case in vertebrates, as well. Although much is understood about how the homeotic genes are regulated in *Drosophila*, for the most part it remains unclear how the Hox genes in vertebrates are regulated. Because these genes are involved in embryonic development, understanding how they are regulated and what regulates them is important.

## Literature Review

### *Embryo patterning models*

In *Drosophila*, embryonic development begins when the basic body plan is laid down, whereas in vertebrates, embryonic development begins before a body plan is established. Patterning of the basic body plan has been best elucidated in *Drosophila*. Five sets of developmental genes are involved in a hierarchy of control initiated by the maternal-effect genes. Maternal-effect gene products, such as *bicoid* and *dorsal*, are laid down in the egg as concentration gradients. During oogenesis these gradients of proteins set up the anterior-posterior (A-P) and dorsal-ventral (D-V) axes, respectively, by acting as positive or negative regulators of the gap genes (Lawrence and Morata, 1994). The gap genes are also spatially expressed as gradients to define the A-P and D-V axes as well as to provide information to the pair-rule genes for their expression (Lawrence and Morata, 1994). Pair-rule genes are expressed in 14 identical domains referred to as parasegments, the first compartments of the embryo. Beginning at gastrulation the next set of genes, the segment-polarity genes, are expressed, regulated by the pair-rule genes. These genes act to further subdivide the parasegments into anterior and posterior halves (Lobe and Gruss, 1989). Finally, the homeotic genes, regulated by gap gene products and the pair-rule gene products, are expressed in a spatially-specific manner from the anterior boundary of a parasegment and subsequently spread posteriorly. The identity of each segment/body part is ultimately dependent on the combination of homeotic genes being expressed (Lawrence and Morata, 1994). This cascade of gene expression, beginning with the maternal-effect genes, sequentially divides the embryo into smaller units with the most specific/local effects controlled by the homeotic genes (Pankratz and Jackle, 1990; Pankratz et al., 1990).

Genetic control of patterning in *Drosophila* differs from that in vertebrates. In *Drosophila* the initial axes are set up prior to gastrulation at or during oogenesis whereas in vertebrates the basic body plan is established during gastrulation. Spemann and Mangold (1924) launched studies of patterning with the "organizer experiment" in *Xenopus* which showed that a secondary body axis could be induced in a host embryo when cells of the dorsal lip were transplanted from a donor embryo. In the dorsal lip region of cleavage stage blastomeres is an area termed the Nieuwkoop center which is able to induce Spemann's organizer in overlying cells at the beginning of gastrulation (Davidson, 1993). Spemann's organizer is further subdivided into head and body axis organizers. Signaling molecules from the Nieuwkoop center which induce the head and body axis organizers include bFGF (Kimelman and Kirschner, 1987; Slack et al., 1987; Green et al., 1990), TGFβ members such

as activin (Jones et al., 1992; Hemmati-Brivanlou and Melton, 1992), *wnt* family members (Christian et al., 1991; Sokol et al., 1991; Smith and Harland, 1991), and *noggin* ( and Harland, 1992). *goosecoid* (*gsc*), a homeobox gene, has been shown to be a direct target of activin (Cho et al., 1991; Watabe et al., 1995) and of *wnt* genes (Watabe et al., 1995) in the organizer. Other possible targets of the signaling molecules in the organizer include forkhead-like genes (*XFKH*1, *XFD-1*, and pintallavis), inducible by activins (Ruiz I Altaba and Jessell, 1992), and *Xnot* (von Dassow et al., 1993). One hypothesis is that *gsc* is induced by activin and is responsible for development of the head region while bFGF induces Ant-type hox genes which are responsible for trunk and tail development (Cho et al., 1991). Steinbeisser et al. (1993) presents results that *gsc* is capable of partial induction of a secondary axis and partial rescue anterior structures, whereas *Xwnt-8* can induce a complete secondary axis and rescue of anterior structures. Also, *gsc* function is induced by both activin and *wnt* genes (Watabe et al., 1995). These results indicate that *gsc* may be necessary, but not sufficient, for induction of head structures and requires both *wnt* genes and activin. Evidence for the requirement for FGF in trunk and tail development was elucidated by Amaya et al. (1993) and Fukui and Asashima (1994). Both show that injection of a dominant negative FGF receptor results in defects of the trunk and tail.

It is believed that there is conservation of the regulatory mechanisms for organizer formation and gastrulation within vertebrates (Watabe et al., 1995). In mouse, the A-P axis is defined by the formation and migration of the primitive streak with further patterning due to signals from tissues arising from the streak. The epiblast cells at the anterior of the streak migrate to form the node. These organizing cells are the equivalent of Spemann's organizer, as the node is capable of inducing a secondary axis (Beddington, 1994). Additionally, possible homologs of genes expressed in Spemann's organizer have been observed in this region. These include a conserved homolog of *gsc* (Blum et al., 1992) and a forkhead related gene, HNF3β (Sasaki and Hogan, 1993). It is the expression of these genes which establish the organizing center. Although the signaling molecules involved in node induction are unknown, some of the possibilities include homologs of those involved in induction of Spemann's organizer; FGF, activin, *wnt*-type (Slack and Tannahill, 1992) and *noggin* (Smith and Harland, 1992). In mouse, Hox genes are expressed during late primitive streak stage (Krumlauf, 1994). Their expression appears first in the posterior region of the streak and then spreads anteriorly. The expression of the Hox genes and the expression of *gsc*, which is found in the anterior of the streak (Blum et al., 1992), support the hypothesis in *Xenopus* of the Hox genes involvement in trunk and tail development and *gsc* being responsible for head development.

***Homeobox genes***

The homeobox, a 180bp sequence in homeotic and homeobox-containing genes (McGinnis et al., 1984; Scott and Weiner et al., 1984), is a highly conserved 60aa homeodomain. Four helices make up the homeodomain. One of these helices, which is similar in all homeobox genes, is necessary for DNA sequence recognition (DeRobertis et al., 1990). Many homeodomain proteins are known to activate or repress target genes including other homeobox genes. *Drosophila* homeotic (HOM-C) genes are found in two complexes, the Antennapedia complex (ANT-C) and the Bithorax complex (BX-C), on a single chromosome. The homeobox sequence from the *Drosophila Antennapedia* gene was used as a probe to screen for homologous genes which have been found in every organism examined (Lawrence and Morata, 1994). In mouse and human there are four clusters or Hox complexes, each on a different chromosome (Lobe and Gruss, 1989; McGinnis and Krumlauf, 1992; Krumlauf, 1994). Organizational and structural similarities with the *Drosophila* homeotic genes include a temporal and spatial colinearity of gene expression, such that the more 3' genes are generally expressed earlier and have a more anterior boundary of expression and the more 5' genes are expressed later and their anterior boundaries of expression are more posterior (Gaunt et al., 1988; Graham et al., 1989). This creates a hierarchy of homeobox protein function known as phenotypic suppression or posterior prevalence where two genes having overlapping expression domains results in the more posterior gene being dominant (McGinnis and Krumlauf, 1992). Additionally, when the four clusters are lined up with the genes from each cluster having similar anterior boundaries of expression are parallel, the genes between clusters are more homologous to each other than they are with other genes within their cluster (Lobe and Gruss, 1989). These genes are considered subfamilies or paralogous groups.

Although the *Drosophila* homeobox-containing genes are the most extensively characterized, knowledge of their function, expression and regulation may not be useful for characterizing the vertebrate homeobox genes as they may not have the same purposes initially (Davidson, 1991). In the patterning models discussed earlier the homeobox genes in *Drosophila* are utilized to generate the body plan, as opposed to vertebrates which use them later in gastrulation following establishment of the body plan (reviewed in Davidson, 1991 and 1993). Specifically, vertebrate Hox genes are believed to be involved in "translating positional information into identity" along the A-P axis (Boncinelli and Mallamaci, 1995). However, the homeotic genes of *Drosophila* are used in a similar manner as the Hox genes of postembryonic vertebrates (Davidson, 1991). Therefore, Hox genes in *Drosophila* and vertebrates are important for patterning the early embryo, although they are utilized in different manners, and

also for later development. One of the most extensively characterized Hox genes, in terms of its expression and structure, is murine *Hoxa-5*.

***Hoxa-5***

*Hoxa-5* is a homolog of the *Drosophila* homeotic gene *sex combs reduced* (*Scr*). Both are expressed in similar regions of their respective embryos (thoracic region) (*Hoxa-5*, Dony and Gruss, 1987; *Scr*, Riley et al., 1987, LeMotte et al., 1989). Ectopic expression of *Hoxa-5* in *Drosophila* results in transformations that are similar to those observed due to ectopic expression of *Scr* (Zhao et al., 1993). Additionally, hoxa-5 can directly activate a target gene of *Scr*, *forkhead* (*fkh*) in *Drosophila*, indicating that *Hoxa-5* is functionally equivalent to *Scr* (Zhao et al., 1993). Because it is a functional homolog of *Scr*, it is possible that they are also regulated in a similar manner.

A 2.7Kb cDNA of *Hoxa-5* was isolated in a library screen of MB66 cells by Fibi et al. (1988). This cDNA contains an open reading frame of 885bp predicting 270 amino acid (aa) protein containing the homeodomain. Zakany et al. (1988) also isolated a cDNA of *Hoxa-5*. The 1.7Kb clone was isolated from a 12.5 day embryonic mouse spinal cord library and contains an open reading frame of 807bp predicting a 269aa homeodomain protein.

Multiple transcripts are detected from the *Hoxa-5* locus. In an F9 cell line induced with retinoic acid, Fibi et al. (1988) detected 1.9 and 4 Kb transcripts and Murphy et al. (1988) detected 3.1 and 5 Kb transcripts by Northern analysis. The 1.9 and 4 Kb transcripts were also detected in embryonic and adult tissues (Fibi et al., 1988, Zakany et al, 1988, Odenwald et al., 1987) as well as 8 and 9 Kb transcripts (Odenwald et al., 1987). These also were detected by Northern analysis. In these analyses the 1.9 Kb transcript seemed to be the major transcript in midgestational embryos (Fibi et al., 1988, Zakany et al., 1988, Odenwald et al., 1987, Murphy et al., 1988) as well as in adult (Odenwald et al, 1987) with the highest levels being observed in the spinal column (midgestational embryos) (Odenwald et al., 1987, Zakany et al., 1988) or spinal cord (adult) (Odenwald et al., 1987).

RNase protection results identified proximal mRNAs initiated at -74 and -44 and distal transcripts initiated at -196 and upstream of -912 (Zakany et al., 1988). Primer extension results also indicate major proximal start sites at -74 and -44 with minor start sites at -78, -84, -90, -94, 106, -136, -150, -152, and -180 (Murphy et al., 1988, Zakany et al., 1988). The RNase protection results indicate the major transcripts in the midgestational embryo are the distal (Zakany et al., 1988), rather than the proximal as seen by Northern (Fibi et al., 1988, Zakany et al., 1988, Odenwald et al., 1987, Murphy et al., 1988) and primer extension (Zakany et al., 1988; Murphy et al, 1988) analyses. The level of transcripts, detected by RNase protection, varied in the newborn tissues (Zakany et al., 1988). RNA expression was

also detected by *in situ* hybridizations on sections of e8-13 in embryonic ribs, vertebrae and their precursors, lung, stomach, gut, kidney, myelencephalon, and spinal cord (Dony and Gruss, 1987). These structures are all located in the thoracic region of the embryo.

Expression of the hoxa-5 protein has been detected by polyclonal antibody as early as embryonic day 7.5 (e7.5, day of vaginal plug detection is first day of gestation unless otherwise indicated) (Tani et al., 1989). At e 8.5 it was detected throughout the embryo but is progressively restricted in its expression pattern as development proceeds. By e17 it is largely restricted to the nervous system with the strongest expression in the spinal cord and less intensely in the brain (Tani et al., 1989). Weak expression is also observed in the lung (Tani et al., 1989). The expression of protein in the brain does not agree with the *in situ* hybridization experiments performed by Dony and Gruss (1987) as they do not observe any RNA expression in the anterior region of early embryos nor do they observe expression in the spinal cord or lung at e18, only in a small region of the gut and a region of the CNS just posterior to the myelencephalon. In addition, no message is detected in the brain of either the newborn or adult by RNAse protection (Zakany et al., 1988). Therefore, the polyclonal antibody may be detecting homeodomain proteins in addition to hoxa-5.

Multiple cis-acting regulatory elements that are responsible for the spatial-specific expression of *Hoxa-5* have been identified by using the LacZ reporter gene. A 915bp sequence upstream of the proximal promoter was identified to be necessary for LacZ expression in the brachial region of the spinal cord (cervical metamere 4 to thoracic metamere 2) in e11-e13 (day of vaginal plug detection is day 0.5 of gestation) transgenic mice (Zakany et al. 1988). The transgene RNA is expressed in the same regions of the spinal cord in e11.5 and the spinal cord and lung of newborn mice as the endogenous RNA with the exception of the newborn kidney (Zakany et al., 1988). Therefore, a transgene expressed from the proximal promoter of *Hoxa-5* only expresses in a subset of the endogenous RNA pattern. This 915bp upstream sequence necessary for this subset of *Hoxa-5* expression was further narrowed to 604bp of sequence having enhancer qualities (Tuggle et al., 1990). Because this 604bp enhancer expresses LacZ in only a subset of the endogenous pattern, it may represent the expression pattern from the proximal promoter only. Furthermore, additional elements may be necessary to recapitulate the entire endogenous pattern. Larochelle et al. (unpublished) has identified 11.1Kb of genomic DNA, including all of *Hoxa-5* and part of the *Hoxa-6* coding sequence, which is able to recapitulate more of the endogenous *Hoxa-5* RNA pattern in a LacZ construct. Therefore, clustering may be important for regulation of these genes because the genes within the cluster may share cis regulatory elements. This has been observed within the Hoxd cluster by moving genes around in the cluster (van der Hoeven et al., 1996) and also with other Hox genes

(*Hoxc-8*, Bieberich et al., 1990; *Hoxb-4*, Whiting et al., 1991; *Hoxb-3*, Sham et al., 1992; *Hoxb-7*, Vogels et al., 1993;). Complete or close recapitulation of the endogenous pattern has only been achieved using the LacZ reporter gene with several Hox genes, *Hoxb-4* (Whiting et al., 1991); *Hoxa-7* (Puschel et al., 1990, Puschel et al., 1991); *Hoxb-1* (Marshall et al., 1992, Studer et al., 1994, Marshall et al., 1994); and *Hoxc-8* (Bieberich et al., 1990), and these constructs have included large regions of genomic DNA.

Functional regions of the hoxa-5 protein have been identified (Zhao et al., 1996). The homeodomain was found to be needed for targeting to the nucleus and for DNA-binding activity. Transcriptional activity and transformation ability lies within the N-terminal 39 amino acids and a YPWM motif near the homeodomain was necessary for biological activity, specifically for protein-protein interactions.

*Hoxa-5* function has been studied by knocking out the function of the endogenous protein. This was accomplished by inserting the neomycin gene into the homeobox of *Hoxa-5* which truncates the hoxa-5 protein such that any protein translated does not contain the DNA binding region of the protein (Jeannotte et al., 1993). Approximately 50% of the homozygous transgenic mice die between birth and weaning while heterozygous and surviving transgenic mice homozygous for the interrupted gene are indistinguishable from their wild type littermates (Jeannotte et al., 1993). No gross organ defects were observed that could account for the deaths of the homozygous animals. The only observable phenotypes were skeletal abnormalities in the cervical and thoracic regions. These abnormalities included both anterior and posterior homeotic transformations of the vertebral column. Both anterior and posterior transformations are also seen with a knockout of *scr* in *Drosophila*, a homolog of *Hoxa-5* (Wakimoto and Kaufman, 1981). Because both anterior and posterior transformations were observed, Jeannotte et al. (1993) believe it is the level of protein expressed that is important for the correct function of hoxa-5. After backcrossing *Hoxa-5-/-* transgenic mice into several inbred strains, the percent of surviving homozygous transgenic mice dropped as low as 12% (Aubin et al., unpublished). Upon closer inspection of *Hoxa-5-/-* pups that died, it was discovered that they exhibited laryngotracheal defects and lung immaturity (Aubin, et al., unpublished). Furthermore, these transgenic mice had altered expression of several genes involved in lung morphogenesis and the regulation of expression of surfactant proteins: TTF-1, HNF3α, and HNF3β (Aubin et al., unpublished). HNF3α and β are homologs of *fkh* which is a downstream target of *Scr* (Panzer et al., 1992) and can be activated by hoxa-5 in *Drosophila* (Zhao et al., 1993). Therefore, HNF3α and β may be direct/indirect targets for regulation by hoxa-5.

**References**

Amaya, E., Stein, P.A., Musci, T.J., and Kirschner, M.W. (1993). FGF signaling in the early specification of mesoderm in *Xenopus*. Development 118:477-87.

Beddington, R.S.P. (1994). Induction of a second neural axis by the mouse node. Development 120:613-620.

Bieberich, C.J., Utset, M.F., Awgulewitsch, A. and Ruddle, F.H. (1990). Evidence for positive and negative regulation of the Hox-3.1 gene. Proc. Natl. Acad. Sci. USA 87:8462-8466.

Biggin, M.D. and Tjian, R. (1989). Transcription factors and the control of *Drosophila* development. TIG 5:377-383.

Blum, M., Gaunt, S.J., Cho, K.W.Y., Steinbeisser H., Blumberg, B., Bittner, D., and DeRobertis, E.M. (1992) Gastrulation in the mouse: the role of the homeobox gene *goosecoid.* Cell 69:1097-1106.

Boncinelli, E. and Mallamaci, A. (1995). Homeobox genes in vertebrate gastrulation. Current Opin. in Gen. and Dev. 5:619-627.

Cho, K.W.Y., Blumberg, B., Steinbeisser, H., and DeRobertis, E.M. (1991). Molecular nature of Spemann's organizer: the role of the *Xenopus* Homeobox gene *goosecoid.* Cell 67:1111-1120.

Christian, J.L., McMahon, J.A., McMahon, A.P. and Moon, R.T. (1991). *Xwnt*-8, a *Xenopus Wnt*-1/*int*-1-related gene responsive to mesoderm-inducing factors, may play a role in ventral mesodermal patterning during embryogenesis. Development 111:1045-1055.

Davidson, E.H. (1991). Spatial mechanisms of gene regulation in metazoan embryos. Development 113:1-26.

Davidson, E.H. (1993). Later embryogenesis: regulatory circuitry in morphogenetic fields. Development 118:665-690.

DeRobertis, E.M., Guillermo, O. and Wright, C.V.E. (1990). Homeobox genes and the vertebrate body plan. Scientific Am. July, 46-52.

Dony, C. and Gruss, P. (1987). Specific expression of the *Hox 1.3* homeo box gene in murine embryonic structures originating from or induced by the mesoderm. EMBO J. 6:2965-2975.

Fibi, M., Zink, B., Kessel, M., Colberg-Poley, A.M., Labeit, S., Lehrach, H. and Gruss, P. (1988). Coding sequence and expression of the homeobox gene *Hox 1.3.* Development 102:349-359.

Fukui, A. and Asashima, M. (1994). Control of cell differentiation and morphogenesis in amphibian development. Int. J. Dev. Biol. 38:257-266.

Gaunt, S.J., Sharpe, P.T. and Duboule, D. (1988). Spatially restricted domains of homeo-gene transcripts in mouse embryos: relation to a segmented body plan. Development 104 Suppl, 169-179.

Graham, A., Papalopulu, N. and Krumlauf, R. (1989). The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. Cell 57:367-378.

Green, J.B.A., Howes, G., Symes, K., Cooke, J. and Smith, J.C. (1990). The biological effects of XTC-MIF: quantitative comparison with *Xenopus* bFGF. Development 108:173-183.

Hemmati-Brivanlou, A. and Melton, D.A. (1992). A truncated activin receptor inhibits mesoderm induction and formation of axial structures in *Xenopus* embryos. Nature 359:609-614.

Jeannotte, L., Lemieux, M., Charron, J., Poirier, F. and Robertson, E.J. (1993). Specification of axial identity in the mouse: role of the *Hoxa-5* (*Hox1.3*) gene. Genes and Dev. 7:2085-2096.

Jones, N. (1990). Transcriptional regulation by dimerization: Two sides to an incestuous relationship. Cell 61:9-11.

Jones, C.M., Lyons, K.M., Lapan, P.M., Wright, C.V.E. and Hogan, B.L.M. (1992). DVR-4 (bone morphogenetic protein-4) as a posterior-ventralizing factor in *Xenopus* mesoderm induction. Development 115:639-647.

Kimelman, D. and Kirschner, M. (1987). Synergistic induction of mesoderm by FGF and TGF-beta and the identification of an mRNA coding for FGF in the early *Xenopus* embryo. Cell 51:869-877.

Krumlauf, R. (1994). Hox genes in vertebrate development. Cell 70:191-201.

Lamb, P. and McKnight, S.L. (1991). Diversity and specificity in transcriptional regulation: The benefits of heterotypic dimerization. Trends Biochem. Sci. 16:417-422.

Lawrence, P.A., and G. Morata (1994). Homeobox genes: Their function in *Drosophila* segmentation and pattern formation. Cell 78:181-189.

LeMotte, P., Kuroiwa, A., Fessler, L. and Gehring, W. (1989). The homeotic gene *Sex combs reduced* of *Drosophila*: gene structure and embryonic expression. EMBO J. 8, 219-227.

Lobe, C.G., and P. Gruss (1989). Mouse versions of fly developmental control genes: Legitimate or illegitimate relatives? The New Biologist 1:9-18.

Marshall, H., Nonchev, S., Sham, M.H., Muchamore, I., Lumsden, A. and Krumlauf, R. (1992). Retinoic acid alters hindbrain Hox code and induces transformation of rhombomeres 2/3 into a 4/5 identity. Nature 360:737-741.

Marshall, H., Studer, S., Popperl, H., Aparicio, S., Kuroiwa, A., Brenner, S. and Krumlauf, R. (1994). A conserved retinoic acid response element required for early expression of the homeobox gene *Hoxb-1*. Nature 370:567-571.

McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A. and Gehring, W.J. (1984). A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and bithorax complexes. Nature 308:428-433.

McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and axial patterning. Cell 68:283-302.

Mitchell, P.J. and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245:371-378.

Murphy, S.P., Garbern, J., Odenwald, W.F., Lazzarini, R.A. and Linney, E. (1988). Differential expression of the homeobox gene *Hox-1.3* in F9 embryonal carcinoma cells. Dev. Bio. 85:5587-5591.

Odenwald, W.F., Taylor, C.F., Palmer-Hill, F.J., Friedrich, V., Jr., Tani, M. and Lazzarini, R.A. (1987). Expression of a homeo domain protein in noncontact-inhibited cultured cells and postmitotic neurons. Genes and Dev. 1:482-496.

Pabo, C.O. and Sauer, R.T. (1992). Transcription factors: Structural families and principles of DNA recognition. Annu. Rev. Biochem. 61:1053-1095.

Pankratz, M., and Jackle (1990). Making stripes in the *Drosophila* embryo. TIG 6:287-292.

Pankratz, M., E. Seifert, N. Gerwin, B. Billi, U. Nauber, and H. Jackle (1990). Gradients of Kruppel and knirps gene products direct pair-rule gene stripe patterning in the posterior region of the *Drosophila* embryo. Cell 61:309-317.

Panzer, S., Weigel, D. and Beckendorf, S.K. (1992). Organogenesis in *Drosophila melanogaster*: Embryonic salivary gland determination is controlled by homeotic and dorsoventral patterning genes. Development 114:49-57.

Puschel, A.W., Balling, R. and Gruss, P. (1990). Position-specific activity of the *Hox 1.1* promoter in transgenic mice. Development 108:435-442.

Puschel, A.W., Balling, R. and Gruss, P. (1991). Separate elements cause lineage restriction and specify boundaries of *Hox-1.1* expression. Development 112:279-287.

Riley, P.D., Carroll, S.B. and Scott, M.P. (1987). The expression and regulation of Sex combs reduced protein in *Drosophila* embryos. Genes and Dev. 1:716-730.

Ruiz I Altaba, A. and Jessell, T.M. (1992). Pintallavis, a gene expressed in the organizer and midline cells of frog embryos: involvement in the development of the neural axis. Development 116:81-93.

Sasaki, H., and Hogan, B.L.M. (1993). Differential expression of multiple fork head related genes during gastrulation and axial patterning in the mouse embryo. Development 118:47-59.

Scott, M.P. and Weiner, A.J. (1984). Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax and fushi tarazu loci of *Drosophila*. Proc. Natl. Acad. Sci, USA 81:4115-4119.

Sham, M. H., Vesque, C., Nonchev, S., Marshall, H., Frain, M., Gupta, R. D., Whiting, J., Wilkinson, D., Charnay, P. and Krumlauf, R. (1993). The zinc finger gene *Krox20* regulates *HoxB2* (*Hox2.8*) during hindbrain segmentation. Cell 72:183-196.

Slack, J.M.W., Darlington, B.G., Heath, J.K. and Godsave, S.F. (1987). Mesoderm induction in early *Xenopus* embryos by heparin-binding growth factors. Nature 326:197-200.

Slack, J.M.W., and Tannahill, D. (1992). Mechanism of antero posterior axis specification in vertebrates. Lessons from the amphibians. Development 114:285-302.

Smith, W.C. and Harland, R.M. (1991). Injected X*wnt*-8 RNA acts early in *Xenopus* embryos to promote formation of a vegetal dorsalizing center. Cell 70:829-840.

Smith, W.C. and Harland, R.M. (1992). Expression cloning of *noggin*, a new dorsalizing factor localized to the Spemann Organizer in *Xenopus* embryos. Cell 70:829-840.

Sokol, S., Christian, J.L, Moon, R.T. and Melton, D.A. (1991). Injected *Wnt* RNA induces a complete body axis in *Xenopus* embryos. Cell 67:741-752.

Spemann, H., and Mangold, H. (1924). Uber Induktion von embryonalanlagen durch implantation artfremder organisatoren. Roux' Arch. f. Entw. Mech. 100:599-638.

Steinbeisser, H., DeRobertis, E.M., Ku, M., Kessler, D.S. and Melton, D.A. (1993). *Xenopus* axis formation: Induction of goosecoid by injected X*wnt*-8 and activin mRNAs. Development 118:499-507.

Studer, M., Popperl, H. and Marshall, H. (1994). Role of a conserved retinoic acid response element in rhombomere restriction of *Hoxb-1*. Science 265:1728-1732.

Tani, M., Odenwald, W.F., Lazzarini, R.A. and Friedrich, V.L., Jr. (1989). Progressive restriction in the distribution of the *Hox-1.3* homeodomain protein during embryogenesis. J. Neuro. Res. 24:457-469.

Tuggle, C.K., Zakany, J., Cianetti, L., Peschle, C. and Nguyen-Huu, M.C. (1990). Region-specific enhancers near two mammalian homeo box genes define adjacent rostrocaudal domains in the central nervous system. Genes and Dev. 4:180-189.

van der Hoeven, F., Zakany, J. and Duboule, D. (1996). Gene transpositions in the HoxD complex reveal a hierarchy of regulatory controls. Cell 85:1025-1035.

Vogels, R., Charite, J., deGraffe, W. and DesChamps, J. (1993). Proximal cis-acting elements cooperate to set *Hoxb-7* (*Hox-2.3*) expression boundaries in transgenic mice. Development 118:71-82.

von Dassow, G., Schmidt, J. and Kimelin, D. (1993). Induction of the *Xenopus* organizer: expression and regulation of *Xnot*, a novel FGF and activin-inducible homeobox gene. Genes and Dev. 7;355-366.

Wakimoto, B.T and Kaufman, T.C. (1981). Analysis of larval segmentation in lethal genotypes associated with the antennapedia gene complex in *Drosophila melanogaster*. Dev. Bio. 81:51-64.

Watabe, T., Kim, S., Candia, A., Rothbacher, U., Hashimoto, C., Inoue, K., and Cho, K.W.Y. (1995). Molecular mechanisms of Spemann's organizer formation: conserved growth factor synergy between *Xenopus* and mouse. Genes and Dev. 9:3038-3050.

Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P.W., Scott, D. and Alleman, R.K. (1991). Multiple spatially specific enhancers are required to reconstruct the pattern of *Hox 2.6* gene expression. Genes and Dev. 5:2048-2059.

Zakany, J., Tuggle, C.K., Patel, M.D. and Nguyen-Huu, C. (1988). Spatial regulation of homeobox gene fusions in the embryonic central nervous system of transgenic mice. Neuron 1:679-691.

Zhao, J.J., Lazzarini, R.A. and Pick, L. (1993). The mouse *Hox-1.3* gene is functionally equivalent to the *Drosophila* Sex combs reduced gene. Genes and Dev. 7:343-354.

Zhao, J.J., Lazzarini, R.A. and Pick, L. (1996). Functional dissection of the mouse *Hoxa-5* gene. EMBO J 15:1313-1322.

# CIS-ACTING ELEMENTS AND TRANS-ACTING FACTORS REGULATING *HOXA-5*: A GRADIENT OF BINDING ACTIVITY TO A BRACHIAL SPINAL CORD ELEMENT

A paper submitted to the journal Developmental Biology

Tamara Nowling, Wen Zhou, Chi Nguyen-Huu, and Christopher K. Tuggle

## ABSTRACT

The *Hox* homeobox genes cooperate in providing positional information needed for spatial and temporal patterning of the vertebrate body axis. However, the biological mechanisms behind spatial *Hox* expression are largely unknown. In transgenic mice, gene fusions between *Hoxa-5* (previously called *Hox-1.3*) 5' flanking regions and lacZ show tissue- and time-specific expression in the brachial spinal cord in day 11-13 embryos. This spatially-specific expression is directed by a 604 bp regulatory region with enhancer properties. Fine-detail mapping of this enhancer has identified several elements involved in region-specific expression. A brachial spinal cord element, is required for expression in the brachial spinal cord. Factors in embryonic day 12.5 nuclear extracts bind this element in electrophoretic mobility shift assays (EMSA) and protect three regions from DNaseI digestion. All three sites contain an AAATAA sequence and mutations at these sites decrease or abolish binding. Furthermore, this element binds specific individual embryonic proteins on a protein blot. The binding appears as a gradient along the A-P axis with two- to three-fold higher levels observed in extracts from anterior regions than from posterior regions. In parallel with the EMSA, the proteins on the protein blot also show decreased binding to probes with mutations at the AAATAA sites. This brachial spinal cord element and its binding proteins are likely to be involved in spatial expression of *Hoxa-5* during development.

## INTRODUCTION

The genetically defined cascade of homeobox gene regulation in *Drosophila*, which ultimately acts to define the components of the embryo, is an excellent model of how spatial expression of genes is regulated to result in segment identification. This cascade of gene regulation begins with the maternal and gap genes setting up gradients of positional information that regulate the pair-rule genes, which, in turn, act to establish the initial spatial expression of the homeobox-containing homeotic genes (reviewed in Lawrence and Morata, 1994). The homeobox is a 60 amino acid binding domain conserved in many regulatory genes, first discovered in *Drosophila* (McGinnis et al., 1984, and Scott and Weiner, 1984). *Drosophila* homeobox-containing genes are the most extensively characterized in terms of function,

expression, and regulation. A large number of mammalian genes with sequence similarity to the *Drosophila* homeotic genes in the Antennapedia and Bithorax complexes have been identified. These *Hox* genes are expressed in many cell types, and their expression domains depend on the tissue/structure and the developmental stage in which they are expressed (Krumlauf, 1994), yet little is understood as to how spatial *Hox* regulation is achieved.

In *Drosophila*, homeotic genes exhibit spatial-specific expression patterns controlled by multiple positive and negative cis-regulatory elements which are responsible for different aspects of their expression patterns (Kennison, 1993; Qian et al., 1991; and Gindhart et al., 1995) . Further, two types of trans-acting factors are involved in regulating the homeotic genes in *Drosophila*: those required for initial patterns in the early embryo (gap and pair-rule genes; Irish et al., 1989, White and Lehmann, 1986) and those required for maintenance of the patterns (trx genes; Kennison and Tamkun, 1992, Pc genes; Paro, 1990). Although some mammalian homologs of the *Drosophila* homeotic regulators have been found (*M33*, Pearce et al., 1992; *bmi-1*, van der Lugt et al., 1994; *brg1*, Randazzo et al., 1994; *Pax*, Dressler et al., 1988), none have been shown to directly regulate *Hox* gene expression. Because it has been suggested that vertebrates may not have similar mechanisms for regulating the *Hox* genes (Krumlauf, 1994), an alternative to searching for direct homologs of genes in the *Drosophila* cascade needs to be employed. Some transcription factors that have been identified as regulators of *Hox* genes include the retinoic acid receptors (Moroni et al., 1993, Popperl and Featherstone, 1993, and Langston and Gudas, 1992), Krox20 (Sham et al., 1993), and GATA-1 in an erythroid cell line (Vieille-Grosjean and Huber, 1995). Also, Hoxb-1 has been shown by Popperl et al. (1995) to autoregulate *Hoxb-1* but requires cofactors. One of these cofactors may be mouse pbx, as its *Drosophila* homolog, exd, a regulator of homeotic genes, can bind cooperatively with Hoxb-1 to pbx consensus sites in *Hoxb-1* in vitro (Popperl et al., 1995).

To identify regulators of vertebrate *Hox* genes, cis-acting regions required for the correct spatial expression of these genes need to be identified. In mouse, cis-regulatory flanking regions of several *Hox* genes are capable of reproducing all or part of the expression pattern of their endogenous genes. Multiple subelements important for different aspects of their expression patterns have been identified in several *Hox* genes (Zakany et al., 1988; Bieberich et al., 1990; Tuggle et al., 1990; Puschel et al., 1991; Whiting et al., 1991; Gerard et al., 1993; Vogels et al., 1993; Brown and Taylor, 1994; Gutman et al., 1994; Shashikant, et al., 1995; Charite et al., 1995); however, the DNA fragments needed to replicate the pattern for most *Hox* genes are very large and few specific regulatory sites bound by trans-acting factors have been identified.

Previous studies of *Hoxa-5* identified a defined, relatively compact cis-regulatory region of 604 bp capable of expressing a subset of its endogenous pattern, the brachial region of the spinal cord, cervical metamere 4 to thoracic metamere 2 (C4-T2) (Tuggle et al., 1990). While the hoxa-5 protein has been shown to bind to the *Hoxa-5* promoter region (Odenwald et al., 1989), binding is outside of the 604 bp enhancer region. A comparison of the 604 bp cis-acting regulatory region of *Hoxa-5* to *Drosophila* and mouse sequences found no similarities to known regulatory sequences, other than several copies of the minimal homeobox binding motif TAAT. Irrespective of whether the 604 bp enhancer faithfully represents the proximal promoter expression pattern, this defined and compact DNA region can be used to study the control of spatial-specific gene expression during development.

By identifying and characterizing trans-acting factors that interact with cis-acting regulatory elements of the *Hoxa-5* enhancer, we hope to develop a model of spatial *Hox* expression. In our current studies, deletion and addition analysis in transgenic mice were used to identify several positive and negative subelements within *Hoxa-5* 5' flanking sequence responsible for different aspects of its region-specific expression in the brachial spinal cord (BSC). Trans-acting factors in nuclear extracts (NE) from midgestational embryos were identified in electrophoretic mobility shift assays (EMSA) that exhibit sequence-specific binding to one of these regulatory elements. We show protection of three sites within a 50 bp region of the BSC element by proteins in the same NE. All three sites contain the sequence AAATAA. Probes containing these sites are bound by renatured proteins separated by SDS-PAGE; intriguingly, these factors exhibit a gradient of binding activity along the A-P axis of the mouse. Furthermore, when these AAATAA sites are mutated, there is a decrease in binding activity in both the EMSA and the protein blot.

## MATERIALS AND METHODS

**Transgene Construction** (In all cases, NotI released the construct for microinjection.)

5' Deletion Analysis: To delete the 5' 292 bp of the enhancer, the EcoRV-XhoI (292-604) fragment was isolated from pCKT32-47 (Tuggle et al. 1990) and ligated back into pCKT32-47 digested with SalI (filled-in) and XhoI. For additional 5' deletions, plasmid pCKT13 was produced by BglII/BamHI digestion of p*Hox1.3*/lacZ/SV40 (Zakany et al., 1988) and ligation into pGEM2 digested with BamHI. BAL31 digestion of pCKT13 digested with SalI was performed. Resulting products were released from the plasmid by digestion with BamHI and cloned into pCKT17-2 (Tuggle et al., 1990) via the BamHI and SmaI sites. Sequencing identified various deletions of the 5' side of the BglII-XhoI enhancer region.

3' Deletion Analysis: To delete the DdeI to XhoI (404-604) and the EcoRV-XhoI (292-604) region, the BglI-XhoI fragment was isolated from pCKT13 by SalI and XhoI digestion

and fragment isolation. Subsequent digestion by either DdeI or EcoRV was followed by Klenow fragment enzyme treatment and ligation into pCKT20 digested with XhoI (filled in) to produce plasmids containing construct 8 and 9. pCKT20 is a BAL31 digestion product of pCKT13 as described above. To produce additional deletions at the 3' end of the enhancer, PCR primers were designed to delete 49 bp (5'-CCCTCGAGCAAGGGCCGGGGTCGAAT), or 98 bp (5'-CCCTCGAGCATTTCCCTCGCAGTTCC) from the 3' side of the enhancer. Using a common 5' primer starting at 185 bp 3' to the BglII site (5'-GATCCTGTCCTTCATGCGTTCACAAAAACAGAGCCGTAAT) the PCR was used to generate DNA fragments of interest. XhoI digestion of the PCR products and ligation into pCKT20 via SmaI and XhoI completed the cloning. Sequencing of plasmids was done to ascertain enhancer sequence was correct and expected deletion had occurred.

Addition Analysis: Construct 10: The BglII to XhoI mouse *Hoxa-5* fragment was filled in and inserted into the EcoRI site at the 5' end of the human D4 sequence by partial EcoRI digestion and fill in with Klenow enzyme of plasmid p5.1/1.3/lacZ (Tuggle et al., 1990). Constructs 11 and 12: The isolated 604 bp *Hoxa-5* fragment was digested with EcoRV. The resulting two fragments 1-292 (Construct 11) and 292-604 (Construct 12) were separately inserted into p5.1/1.3/lacZ, as for construct 10. Construct 13: The isolated 604 bp *Hoxa-5* fragment was digested with DdeI and filled in with Klenow. The resulting fragment 1-404 was inserted into the EcoRI site of 5.1/1.3/lacZ as above. Constructs 14 and 15: Oligonucleotides for both strands covering the 292 to 367 region and for the 333-407 region were designed with BamHI cohesive ends. Annealed oligos were digested with BamHI and ligated into BamHI-digested p5.1/1.3/lacZ or into pCKT17-2 digested with BamHI to generate pCKT39-12 and pCKT38-6, respectively.

**Transgenic Mice Production and Analysis**

All transgenic mice were produced by microinjection of linearized DNA fragments into C57B6/CBA F2 embryos as described (Hogan et al., 1986). DNA fragments prepared by CsCl centrifugation or by low melt agarose/phenol extraction followed by dialysis gave similar efficiencies. All potential transgenic conceptuses were dissected out at day 12.5 of gestation and analyzed for lacZ expression (embryo) and transgene presence (placenta) as previously described (Zakany et al., 1988). The morning of vaginal plug detection was designated as 0.5 day of gestation.

**Electrophoretic Mobility Shift Assay**

Crude nuclear extract (NE) (Roy et al, 1991) was prepared from either whole embryonic 9.5-18.5 day postconception (p.c.) mice or embryo sections (1-4) along the A-P axis of embryonic 12.5 day p.c. (e12.5) mice. Regions 1 through 4 include: (1) head, to the

lower jaw; (2) upper cervical, from lower jaw to just anterior of the forelimbs; (3) brachial, from anterior of forelimbs to just posterior to the forelimbs; and (4) the rest of the torso. Where indicated, the crude NE was further purified by step gradient fractionation using HiTrap Heparin columns (Pharmacia BioTech). Protein concentrations were determined either by a protein spectrophotometric assay, measuring absorbance at 230 and 260nm (Kalb and Bernlohr, 1977), or by the Bradford microassay.

The 75 bp fragments used for the brachial spinal cord (BSC) EMSA were isolated from pCKT38-6 and pCKT39-12 by BamHI digests and end-labeled using the Klenow reaction. Binding reactions and gel shifts were performed as described in Roy et al. (1991) with the following modifications: reactions were incubated at room temperature for 20 minutes in the presence of polyd(A:C)d(G:T) and run on a 5%, 29:1 acrylamide:bisacrylamide, 0.25X TBE nondenaturing gel. Approximately 0.5-1 ng of probe and 100-200 ng of NE were used per reaction for the mutant EMSA analyses and the same amount of probe and 12.5-25 ng was used for the EMSA analyses across sections of the embryo. The lower amount of extract was necessary to observe binding activity within the linear range for quantitation purposes. For competition studies, specific or nonspecific competitor DNA was added to the reactions before addition of the labeled probe. After electrophoresis, the gel was dried and autoradiography performed.

Molecular weights of shifted complexes were estimated following the protocol of Orchard and May (1993) using molecular weight markers from Sigma. Specific shifted bands were quantitated by phosphorimaging. The percentage of DNA bound for each fragment was calculated by dividing the amount of radioactivity in the shifted band by the total amount of radioactivity in the lane (shifted band plus free band). Quantitation data include at least triplicate assays in all analyses.

**Footprinting**

Probes used in the footprinting assays were generated by filling in the Sau3AI site of the 160 bp EcoRV/Sau3AI (E/S) restriction fragment and the TaqI site of the 140 bp TaqI/HinfI (T/H) restriction fragment, both from pBglXho (Tuggle et al., 1990), with $\alpha$-$^{32}$P dCTP using the Klenow reaction. The E/S and T/H fragments were used to footprint the top and bottom strands, respectively. The DNaseI protection footprinting was performed as described by Larkin (1993) using the described probes and fractionated e12.5 day extract. Binding reactions were carried out as in the EMSA analyses before treatment with DNaseI. The treated binding reactions were denatured and run on a 7%, 19:1 acrylamide:bisacrylamide, 1X TBE, 8M urea sequencing gel. After electrophoresis, the gel was fixed, dried, and autoradiography performed.

**PCR Production of Mutant Regulatory Element Probes**

Protected sites identified by footprinting analysis were mutated using the polymerase chain reaction (PCR). Primers were designed with changes at four bases within each protected region, (A, B, and C) and between regions B and C as a control. All primers were synthesized at the Sequencing and Synthesis Facility at Iowa State University. Primers include:
primer 1, 5'TGGTGACTTAGAATcgggTACAACAAC3';
primer 2, 5'ACTTAGAATTATTTACAACAACTTcgggTCCCCGG3';
primer 3, 5'GGAGGTGGTGACTTAGAATcgggTACAACAACTTcgggTCCCC3';
primer 4, 5'ACCCCAACCTCTACACAAAAgcccAGAGGGG3';
primer 5, 5'TGGCAAACCGACCCCAACCTCTA3';
primer 6, 5'CGGGGGAGGTGGTGACTTAGAATTA3'; and
primer 7, 5'GTGGTGACTTAGAATTATTTgacgCAACTT3'. Primers 5 and 6 generate a wild type (WT) probe, primers 4 and 6 generate a probe with a mutant A (MutA) site, primers 5 and 2 a mutant B (MutB) site, primers 5 and 1 a mutant C (MutC) site, primers 4 and 2 generate an AB double mutant (MutAB), primers 4 and 1 generate an AC double mutant (MutAC), primers 5 and 3 generate a BC double mutant (MutBC), primers 4 and 3 generate an ABC triple mutant (MutABC), and primers 5 and 7 generate a mutant at a site between B and C as a control mutation (MutCtrl). For EMSA, one primer from each primer set was end labeled with $\gamma$-$^{32}$P ATP and then used in the PCR reactions. For southwesterns, the probes were labeled by incorporation of $\alpha$-$^{32}$P dCTP by PCR. For compatibility purposes in the PCR reaction, primer design software dictated slight differences in product size. With the exception of products generated using primer 4, these products contain all of the original 3' BSC (333-404) sequence with the addition of only 6-14 bp of flanking sequence. Based on our results with the control product (MutCtrl), this additional sequence does not affect binding. Products generated using primer 4 begin 3 bp within the original 3' BSC sequence on the 5' end; however, these 3 bp are unnecessary for binding because one of these products, MutA, showed little effect on binding. Bands were quantitated by phosphorimaging. Data are based on at least three trials for all probes.

**Protein Blot Analysis**

Southwestern analyses were performed as described previously (Cowell and Hurst, 1993). Equal amounts (25-50 μg) of NE from whole or sections 1 through 4 of e12.5 embryos were separated on a 12% SDS-PAGE gel, transferred to nitrocellulose, renatured, and incubated with probe. Probes included $^{32}$P$\gamma$ATP labeled and ligated 65 bp upper cervical repressor (UCR) fragment and 75 bp 3' (333-404) BSC fragment, or PCR-generated 3' BSC wild type (WT) and mutant probes MutB, MutBC, and MutABC. For the southwestern with

NE from sections, detected bands in each section were quantitated by phosphorimaging and normalized within each section to a nonspecific band (bands that co-migrate on the UCR and 3' BSC probed blots), which varied no more than 16% across the sections of the embryo, as a control for transfer and probe binding efficiency on each blot. Section 4 (having the smallest ratio) for each band in each experiment was set equal to one and the other sections were adjusted accordingly. This was done to more accurately compare experiments because the overall level of radioactivity binding between experiments varied, likely due to differences in washing or specific activity of the probe. High and low molecular weight markers (Biorad) were used for size estimation. For the southwestern with whole NE and the PCR-generated probes, the binding in bands for the mutants was quantitated by phosphorimaging and calculated as a percentage of the WT. Quantitation data are based on triplicate southwesterns that included two separate extract preparations.

## RESULTS

### Identification of Regulatory Elements in the *Hoxa-5* Region-Specific Enhancer

Previous analysis of *Hoxa-5* (*Hox-1.3*) identified a 604 bp region (BglII- XhoI, see Figure 1) which conferred brachial spinal cord (BSC) (C4-T2) expression to a lacZ reporter gene in transgenic embryos (Zakany et al., 1988 and Tuggle et al, 1990). This fragment was active in either orientation, and activated the mouse hsp68 promoter as well, indicating that enhancer-like elements existed in this fragment. Because the lacZ expression pattern was temporally and spatially specific, expressing β-galactosidase from embryonic day 11 to 13 with discrete rostral and caudal boundaries, the *Hoxa-5* BSC enhancer was investigated as a model to dissect region-specific expression. We used deletion analysis to further define elements within the 604 bp region. Figure 1 shows the various fragments injected to generate transgenic embryos and Figure 2 shows day 12.5 embryos displaying representative lacZ expression patterns.

When sequences up to 211 bp 3' to the BglII site were removed, no change was observed; however, when 237 bp were deleted, BSC expression was lost, indicating that the additional 26 bp removed is necessary for BSC expression. In a similar analysis using 3' deletions, several effects were observed. If the 3' 50 bp were deleted (3'Δ50), more anterior lacZ expression was seen, to the C1-C2 level (Figure 2D & E). BSC expression and the T2 caudal limit of expression was maintained, however, indicating there is a negative element(s) specifically regulating the anterior limit within the 555-604 bp region, which we termed the upper cervical repressor (UCR) element. If 100 bp at the 3' end are deleted, the expression pattern changes significantly. BSC expression is lost, while expression appears in the peripheral nervous system (Figure 2F & G). This expression was limited to the same

embryonic rostro-caudal region seen for the 3'Δ50 construct, indicating region-specificity has been maintained whereas cell-type specificity has changed. If 200 bp 3' were removed, no expression was observed, indicating additional elements required for expression exist within the 404-505 region (data not shown).

These results show that several regulatory elements are present within the *Hoxa-5* enhancer and that about 400 bp (211 to 604) is required for BSC expression. To further analyze this region, small fragments of the *Hoxa-5* enhancer were added to another gene fusion (*HOXD4*/lacZ), which is expressed in the same tissue during midgestation, but in a different region of the embryo (previously described as *Hox5.1*/lacZ, Tuggle et al., 1990). By adding *Hoxa-5* elements to the *HOXD4*/lacZ, which expresses lacZ from the myelencephalon to C3, we hoped to eliminate the requirement of non-spatial regulatory elements in the analysis of the BSC enhancer. Placing the entire 604 bp region 5' to the 2832 bp *HOXD4* fragment produced embryonic expression in both regions (myelencephalon to T2, Figure 2H), showing that the two enhancers seem to work in an independent and positive manner. Systematic addition analysis of different fragments of the *Hoxa-5* BSC enhancer region to the *HOXD4* construct identified a small region (292-404) required for BSC expression in the context of the *HOXD4* enhancer (Figure 2L and 2M). Addition of either the 5' (292-367) or 3' (333-404) two-thirds of this region also produced BSC expression, although only the 3' (333-404) sequence produced strong BSC expression (Figures 1 and 2). These results indicate that spatial elements for BSC expression exist within this latter 72 bp region (summarized in Figure 1B).

**Specific Binding of Embryonic Proteins to the *Hoxa-5* BSC Element**

To identify trans-acting factors binding to the cis-regulatory elements described above, electrophoretic mobility shift assay (EMSA) analyses were performed. Either the 5' (292-367) or 3' (333-404) two-thirds of the BSC element were used in a binding assay with embryonic e12.5 NE. In the EMSA several complexes were detected. Competition studies show these DNA-binding proteins are sequence-specific and indicate that the specific binding is contained entirely within the 3' 333-404 sequence (Figure 3). The two strongest specific bands have protein components with molecular weights of 122 and 89 kDa. The level of binding to the 3' BSC element within extracts from different e12.5 anterior-posterior sections (1-4) was quantitated (data not shown). These results indicate a slight increase in the activity of these complexes from head to tail, although only section 1 and 4 binding levels were statistically different using the two-tailed t-test ($p < 0.05$).

**Three Regions of Binding Detected by DNaseI Protection Studies All Contain the Sequence AAATAA**

Sequence comparison of the 604 bp *Hoxa-5* enhancer was performed with known *Drosophila* and mouse transcription factor binding site consensus sequences, including: trithorax, polycomb, paired-box, oct, RAR, pax, as well as intergenic regions of other *Hox* genes. Other than five homeobox motif sequences (TAAT) from 216-521 bp (see Figure 1A), no significant homology was found between the enhancer and any of these binding sites. Therefore, we performed footprinting assays to determine at what sites the factors, observed in the EMSA, were binding. DNaseI protection footprinting experiments were performed with the BSC element and fractionated e12.5 NE. Strong protection of one region, footprint B, and weak protection at two other regions, footprints A and C (Figure 4A and 4B) were observed. All three regions contain the sequence AAATAA (Figure 4C) and all are found within the 3' (333-404) BSC fragment.

**Individual Proteins Exhibit a Gradient of Binding Activity to a *Hoxa-5* Regulatory Element**

Because the sequence-specific DNA:protein complexes observed in the EMSA may contain multiple proteins, we performed an assay to determine if individual factors could bind the 3' BSC element. Equal amounts of NE from embryo sections were subjected to SDS-PAGE, transferred to nitrocellulose, renatured, and probed with either the 3' BSC element or the UCR element. The UCR probe detects three bands and the BSC probe detects six bands (Figure 5A). The three bands on the UCR protein blot co-migrate with three of the bands on the BSC blot, indicating they may be nonspecific binding proteins. The three unique bands on the BSC protein blot were deemed likely to be specific for the BSC element. Their molecular weights were estimated at 38, 36, and 34 kDa.

Interestingly, results from quantitation of these three specific bands normalized to one of the nonspecific bands (Figure 5B) or non-normalized (data not shown), indicate they exist as gradients along the A-P axis. Binding observed in the D band increases nearly three fold from tail to head and binding in the E and F bands increase by approximately 2.5 fold from tail to head. Because the non-specific bands used for normalization varied only 7-16% across sections, we are confident that the 38, 36, and 34 kDa bands are exhibiting a gradient of activity.

**Each Protected AAATAA Element is Necessary for Maximal Binding Activity**

To refine the regulatory sequences recognized by embryonic proteins and to test the significance of the footprinted sites on the BSC element, single (MutA, B, C, and Ctrl), double (MutAB, AC, and BC) and triple (MutABC) mutations were generated (Figure 4C). These

fragments were then used in EMSA analyses (Figure 6A and 6B) and the level of binding for each fragment quantitated (Figure 6C). Mutating just the A site or just the C site has little to no effect on binding. When only the B site is mutated there is an 83% reduction in binding when compared with the wild type fragment. When sites B and C are both mutated there is a further reduction in binding of about 9% over mutant B alone. Although A and C do not have much effect on binding alone an AC double mutation reduces binding by 89%, while an AB double mutation results in a 95% reduction in binding. When all three sites are mutated, binding is reduced to nearly undetectable levels. Mutating sequences between sites B and C (MutCtrl) has very little effect on protein binding.

To correlate the proteins binding in the EMSA with those binding in the protein blot, the WT and B, BC, and ABC mutant probes were used in the southwestern assay with embryonic extracts. Binding in bands D and E (Figure 7A) was quantitated as a percentage of the WT binding, as the binding in band F was too light to be quantitated with PCR-generated probes. There is a 50-75% reduction in binding with the MutB and MutBC probes compared to the WT and a 90% reduction in binding with the MutABC probe (Figure 7B). Thus, mutations in the AAATAA consensus sites, which result in a reduction of binding in the EMSA, also result in a reduction of binding in the protein blot. Furthermore, the levels of reduction binding activity in the protein blot correlates with the EMSA activity seen with each mutation.

## DISCUSSION

### The *Hoxa-5* Enhancer is a Complex Enhancer Containing Multiple Elements for Spatial-Specific Expression.

Several elements responsible for different aspects of *Hoxa-5* spatial-specific expression were identified by *in vivo* analysis of expression from various *Hoxa-5*/lacZ transgenes. 5' deletion analyses of the 604 bp enhancer identified a 26 bp region that is necessary for BSC expression. This 26 bp segment is highly (70%) homologous to a region in the *HOXD4* enhancer, which is also expressed in the spinal cord, thus this 26 bp segment may be a spinal cord-specific element. Two other elements, located at the 3' end of the enhancer, identified by deletion analysis include an UCR element and a peripheral nervous system element. Thus, a large region is required for BSC expression. To further define spatial elements within this region, *Hoxa-5* enhancer sequences were added to another *Hox* construct, *HOXD4*/lacZ, that expresses β-galactosidase in the same tissue in the adjacent upper cervical region. Addition of minimal *Hoxa-5* sequences from this large region showed that within the 333-404 sequence is an element(s) capable of conferring BSC lacZ expression in the context of the *HOXD4* enhancer.

**Are These Regulatory Elements Relevant *in vivo*?**

The *Hoxa-5* regulatory elements show a more posterior rostral boundary of expression than the endogenous *Hoxa-5* mRNA (Zakany et al., 1988). This has been observed with other *Hox* transgenes (Vogels et al., 1993; Gerard et al., 1993; Charite et al., 1995). The more posterior rostral expression boundary is most likely due to missing regulatory elements, although using larger regions of some *Hox* genes(5 Kb of *Hoxa-5*, Zakany et al., unpublished and ~27 Kb of *Hoxb-7*, Vogels et al., 1993) resulted in no change in their anterior expression boundaries. However, in the case of *Hoxa-5*, a repressor element (UCR) was identified, the presence of which is necessary to retain the C4 limit of expression in transgenics. This brachial region may be highly relevant *in vivo* because most of the homeotic effects observed in *Hoxa-5* -/- mice are within the C4-T2 region (Jeannotte et al., 1993).

Additionally, *Hoxa-5* is expressed from at least two promoters, proximal and distal, which generate different transcripts expressed in the spinal cord, prevertebrae, lungs, kidney, and gut (Zakany et al., 1988 and Jeannotte et al., 1993). All of these transcripts are detected simultaneously by the probes used in published *in situ* RNA expression studies (Dony and Gruss, 1987). The proximal promoter has been mapped within the XhoI-SacI region (Figure 1) (Zakany et al., 1988). This promoter directs synthesis of a 1.8 Kb transcript, which in Northern blot analysis of poly A+ mRNA, is the major transcript during embryogenesis (Odenwald et al., 1987, Zakany et al., 1988, and Fibi et al., 1988). In RNase protection experiments using total mRNA, however, the majority of embryonic transcripts are initiated 5' to this promoter and the ratio of proximal to distal transcripts varied among neonatal tissues (Zakany et al., 1988). Northern blot analyses by Jeannotte et al. (1993) indicate the ratio of these transcripts also vary in adult tissues, indicating the two promoters are not entirely co-regulated. Because the proximal promoter may be regulated differently from the other *Hoxa-5* promoter(s), and thus may not parallel the RNA pattern detected by *in situ* hybridization, the BSC pattern may represent the *in vivo* protein expression pattern from the proximal promoter. In any event, as so little is understood about *Hox* regulation, studying this 604 bp enhancer region can be useful in identifying cis- and trans-regulatory elements of *Hoxa-5*.

**The BSC Element Contains Three Protected Regions All Containing the Sequence AAATAA.**

Electrophoretic mobility shift assays (EMSA) identified strong specific binding of embryonic proteins in crude NE to the 3' (333-404) BSC element and weaker binding to the 5' (292-367) BSC element. These results support the lacZ expression studies that showed stronger BSC expression with the 3' element. To directly identify protein binding sites, DNase I protection footprinting experiments were performed. Protection from digestion was observed

for three regions on the *Hoxa-5* BSC element. Interestingly, although one site was clearly more occupied (site B), all three of these regions include an AAATAA sequence. One of these footprinted regions, A, is located at 351-360 and is present on both the 3' and 5' fragments. Regulatory protein binding to site A may be responsible for the weak binding and weak lacZ expression seen with the 5' fragment in the EMSA and the transgenics, respectively. All three footprinted regions are located within the 3' fragment, resulting in the strong binding seen in the EMSA.

To further test whether the AAATAA sequences are important in binding, several point mutations were made at each footprint within this 6 bp sequence within the context of the 3' element, and used in the EMSA. Mutating site B significantly reduces binding whereas mutating only sites A or C has very little effect compared with the wild type fragment. The BC double mutant reduces binding somewhat more than the B mutant alone, the double AB and AC mutants reduce binding significantly relative to the A mutant, and the triple ABC mutant abolishes binding completely. Therefore, the B site is the strongest binding site although the A and C sites are important for binding as seen in the AC mutant, which is strongly affected although site B is intact. We note that within site C there is a potential Hox protein binding site (TAAT) and that site C is clearly protected in the DNaseI analysis. However, the mutation in site C knocks out the putative Hox protein binding site and does not seem to be as important as site B in the context of the single mutations, thus the regulatory protein recognizing site C is not likely to be a Hox protein. It would be interesting to analyze the significance of these protected regions by introducing these mutations in the context of the entire enhancer into transgenic mice.

**Individual Embryonic Proteins Bind to the 3' BSC Element and Show a Gradient of Binding Across the A-P Axis of the Embryo.**

The proteins protecting the BSC element recognize multiple similar core sequences (AAATAA), although additional sequences around this motif are evidently used to discriminate between these sites. To determine if a single protein was recognizing the AAATAA containing BSC element, proteins from embryonic nuclear extracts (NE) were denatured and submitted to SDS-PAGE, renatured, and probed with the 3' element. The protein blot included equal amounts of NE made from different sections of the embryo along the A-P axis. Individual proteins specifically binding to this element could be detected in this southwestern analysis and are approximately 34, 36, and 38 kDa in size. In experiments using probes with mutations in the AAATAA sites, significant loss of binding was observed relative to a wild type probe and binding levels correlated with those observed in the EMSA for the different mutant probes. Based on these mutant BSC studies the individual proteins are binding the same AAATAA sites

as the complex of proteins in the EMSA. Therefore, it is likely the individual proteins are a part of this complex and may multimerize with themselves or each other to give the approximate molecular weights calculated for the protein components (89 and 122 kDa) in the EMSA.

Quantitation of probe binding these blots indicate that binding to these specific individual proteins is highest in the head and decreases along the embryo to a lowest level in the posterior region. Each specifically binding protein on the blot displayed a similar gradient, indicating they may be differently modified forms of the same protein. This is substantiated by the fact that the binding of proteins in both bands D and E is reduced with either a single, double or triple mutant BSC probe. But, EMSA results using the same NE from sections do not exhibit the same gradient of binding activity (data not shown). When performed within the linear range of binding activity, the EMSA show a shallow gradient in the opposite direction, low in the head and higher in the tail. Because shifts seen in the EMSA are likely due to a complex of proteins, a gradient similar to that in the southwestern analysis may not be observed in the EMSA because protein-protein interactions may prevent or accelerate binding to the 3' BSC fragment. Such interactions would not occur in the protein blot analysis, as proteins are separated and fixed to the membrane. The southwestern is, therefore, a more direct measure of any one protein binding individually.

The observed gradient of binding activity to the BSC regulatory element is of particular interest considering spatial expression models that involve gradients (Wolpert, 1989). In *Drosophila*, A-P development is dependent on protein and mRNA gradients, such as *bicoid* in the anterior (Nusslein-Vollhard et al., 1987 and Struhl, 1989) and *caudal* in the posterior (MacDonald and Struhl, 1986). In the mouse, genes in the *Hox B* cluster exhibit nested expression domains in the CNS with respect to the A-P axis (Graham et al., 1989 and Hunt et al., 1991) suggesting the *Hox* genes are expressed in a graded manner. Identifying and characterizing these newly detected *Hoxa-5* BSC-binding proteins may help in understanding some of the molecular mechanisms underlying spatial expression of *Hoxa-5* and other genes during development. Because at least some of the BSC element binding proteins are active as individual proteins, a screen for cDNAs encoding these proteins might be successful.

**Regulation of *Hoxa-5* Spatial-Specific Expression.**

Based on our results we can speculate about the regulation of the spatial-specific expression of *Hoxa-5*/lacZ observed in the spinal cord in e12.5 transgenic embryos. Because the 3' (333-404) BSC element is able to confer BSC expression to another *Hox* enhancer, this element is likely to be a positive regulatory element. The 3' 50 bp UCR element, which when deleted from the enhancer results in more anterior lacZ expression, is likely to be a negative

regulatory element. These two elements may be partly responsible for establishing the anterior limit (C4) of *Hoxa-5*/lacZ expression. An example of this is seen in *Drosophila*, where an enhancer element of the *Ubx* gene contains a repressor site which is responsible for establishing the anterior boundary of *Ubx* (Qian et al., 1991). Proteins specifically binding the UCR negative regulatory element have been detected (T. N. and C. K. T., unpublished data) and could turn off expression in the anterior; those binding the BSC positive regulatory element would turn on expression in the brachial region. We speculate that there must be another regulatory factor binding to the BSC element in the EMSA in addition to those detected on the protein blot. This postulated protein would be binding in a gradient opposite that of the proteins on the protein blot (greater binding in the tail, less in the head). This may explain the slight gradient from tail (high) to head (low) observed in the EMSA.

In the posterior region, the sharp posterior limit of *Hoxa-5*/lacZ expression at the level of T2 could be due to suppression of *Hoxa-5* by more posteriorly expressed *Hox* genes, such as *Hoxa-6*, *Hoxa-7*, and *Hoxa-9*. Posterior prevalence, called phenotypic suppression in *Drosophila*, where two genes having overlapping expression domains results in the more posterior gene being dominant (McGinnis and Krumlauf, 1992), has been observed in the case of *Scr* (*Drosophila* homolog of *Hoxa-5*) expression. BX-C genes, including *Antp*, *Ubx*, and *AbdB*, the *Drosophila* homologs of *Hoxa-6*, *Hoxa-7*, and *Hoxa-9*, respectively, are more posteriorly expressed than the ANT-C *Scr* gene, and have been shown to limit expression of *Scr* in some tissues in *Drosophila* (Pelaz et al., 1993 and Andrew et al., 1994). Both *Scr* and *Hoxa-5* are expressed in similar regions of their respective embryos (thoracic) (*Hoxa-5*, Dony and Gruss, 1987; *Scr*, Riley et al., 1987 and LeMotte et al., 1989) and Zhao et al., (1993) showed that *Hoxa-5* is a functional homolog of *Scr*. Therefore, *Hoxa-5* expression may be regulated in a similar manner as *Scr*. Additional evidence that other Hox proteins and hoxa-5 itself may be important for *Hoxa-5* expression includes, several potential Hox binding sites located within the *Hoxa-5* enhancer and the binding of hoxa-5 to its own promoter (Odenwald et al., 1989). However, when the *Hoxa-5*/lacZ transgene was introduced into *Hoxa-5*-/- mice, the lack of endogenous hoxa-5 protein had no effect on the lacZ expression pattern (L. Jeannotte, personal communication). It would therefore be of interest to determine if Hox proteins bind the *Hoxa-5* BSC enhancer, particularly in the 400-550 region which has several TAAT motifs and is required for spatial expression.

## ACKNOWLEDGMENTS

The authors thank Michelle Aulik and Rosalyn Juergens for assistance with EMSA gels, Jozsef Zakany for histological analysis of construct 7 embryos, Percy Liu for plasmid preparations, and Marit Nilsen-Hamilton, Drena Dobbs, and Lucie Jeannotte for critical

comments on the manuscript. Journal paper no. J-16534 of the Iowa Agric. and Home Econ. Exp. Sta., Project no. 3038.

**REFERENCES**

Andrew, D. J., Horner, M. A., Petitt, M. G., Smolik, S. M. and Scott, M. P. (1994). Setting limits on homeotic function: restraint of *Sex combs reduced* activity by *teashirt* and other homeotic genes. *EMBO J.* **13**, 1132-1144.

Bieberich, C.J., Utset, M.F., Awgulewitsch, A. and Ruddle, F.H. (1990). Evidence for positive and negative regulation of the Hox-3.1 gene. *Proc. Natl. Acad. Sci. USA* **87**, 8462-8466.

Brown, W. M. and Taylor, G. R. (1994). The 5'-sequence of the murine *Hox-B3* (*Hox-2.7*) gene and its intron contain multiple transcription-regulatory elements. *Int. J. Biochem.*. **26**, 1403-1409.

Charite, J., de Graaff, W., Vogels, R., Meijlink, F., and Deschamps, J. (1995). Regulation of the *Hoxb-8* gene: Synergism between multimerized cis-acting elements increases responsiveness to positional information. *Dev. Biol.* **171**, 294-305.

Cowell, I. G. and Hurst, H. C. (1993). Cloning transcription factors from a cDNA expression library. *In* "Transcription Factors: A Practical Approach" (D.S. Latchman, Ed.), pp. 110-112. Oxford University Press, New York.

Dony, C. and Gruss, P. (1987). Specific expression of the *Hox 1.3* homeo box gene in murine embryonic structures originating from or induced by the mesoderm. *EMBO J.* **6**, 2965-2975.

Dressler, G. R., Deutsch, U., Balling, R., Simon, D., Guenet, J. L. and Gruss, P. (1988). Murine genes with homology to *Drosophila* segmentation genes. *Development.* **104** (Suppl.), 181-186.

Fibi, M., Zink, B., Kessel, M., Colberg-Poley, A. M., Labeit, S., Lehrach, H. and Gruss, P. (1988). Coding sequence and expression of the homeobox gene *Hox 1.3*. *Development.* **102**, 349-359.

Gerard, M., Duboule, D. and Zakany, J. (1993). Structure and activity of regulatory elements involved in the activation of the *Hoxd-11* gene during late gastrulation. *EMBO J* **12**, 3539- 3550.

Gindhart, Jr., J. G., King, A. N. and Kaufman, T. C. (1995). Characterization of the cis-regulatory region of the *Drosophila* homeotic gene *Sex combs reduced*. *Genetics* **139**, 781-795.

Graham, A., Papalopulu, N. and Krumlauf, R. (1989). The murine and *Drosophila* homeobox complexes have common features of organization and expression. *Cell* **57**, 367-378.

Gutman, A., Gilthorpe, J., and Rigby, P. W. J. (1994). Multiple positive and negative regulatory elements in the promoter of the mouse homeobox gene *Hoxb-4*. *Mol. Cell. Biol.* **14**, 8143-8154.

Hogan, B. L. M., Costantini, F. and Lacy, E. (1986). Manipulating the mouse embryo: A laboratory manual. Cold Spring Harbor Press, Cold Spring Harbor, New York.

Hunt, P., Gulisano, M., Cook, M., Sham, M.-H., Faiella, A., Wilkinson, D., Boncinelli, E. and Krumlauf, R. (1991). A distinct hox code for the branchial region of the vertebrate head. *Nature* **353**, 861-864.

Irish, V. F., Martinez-Arias, A. and Akam, M. (1989). Spatial regulation of the *Antennapedia* and *Ultrabithorax* homeotic genes during *Drosophila* early development. *EMBO J* **8**, 1527-1537.

Jeannotte, L., Lemieux, M., Charron, J., Poirier, F. and Robertson, E. J. (1993). Specification of axial identity in the mouse: role of the *Hoxa-5* (*Hox 1.3*) gene. *Genes Dev.* **7**, 2085-2096.

Kalb, V. F., Jr. and Bernlohr, R. W. (1977). A new spectrophotometric assay for protein in cell extracts. *Anal. Biochem.* **82**, 362-371.

Kennison, J. A. and Tamkun, J. W. (1992). Trans-regulation of homeotic genes in *Drosophila*. *The New Biologist* **4**, 91-96.

Kennison, J. A. (1993). Transcriptional activation of *Drosophila* homeotic genes from distant regulatory elements. *Trends Genet.* **9**, 75-79.

Krumlauf, R. (1994). *Hox* genes in vertebrate development. *Cell* **78**, 191-201.

Langston, A. W. and Gudas, L. J. (1992). Identification of a retinoic acid responsive enhancer 3' of the murine homeobox gene *Hox-1.6*. *Mech. Dev.* **38**, 217-228.

Larkin, N. D. (1993). Determination of DNA sequences that bind transcription factors by DNA footprinting. *In* "Transcription Factors: A Practical Approach" (D.S. Latchman, Ed.), pp. 32-40. Oxford University Press, New York.

Lawrence, P. A. and Morata, G. (1994). Homeobox Genes: Their function in *Drosophila* segmentation and pattern formation. *Cell* **78**, 181-189.

LeMotte, P., Kuroiwa, A., Fessler, L. and Gehring, W. (1989). The homeotic gene *Sex combs reduced* of *Drosophila*: gene structure and embryonic expression. *EMBO J.* **8**, 219-227.

MacDonald, P. M. and Struhl, G. (1986). A molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern. *Nature* **324**, 537-545.

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. and Gehring, W. J. (1984). A conserved DNA sequence in homeotic genes of the *Drosophila Antennapedia* and *Bithorax* complexes. *Nature* **308**, 428-433.

McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* **68**, 283-302.

Moroni, M. C., Vigano, M. A. and Mavilio, F. (1993). Regulation of the human *HoxD4* gene by retinoids. *Mech. Dev.* **44**, 139-154.

Nusslein-Vollhard, C., Frohnhofer, H. G. and Lehman, R. (1987). Determination of anteroposterior polarity in *Drosophila*. *Science* **238**, 1675-1681.

Odenwald, W. F., Taylor, C. F., Palmer-Hill, F. J., Friedrich, Jr., V., Tani, M. and Lazzarini, R. A. (1987). Expression of a homeo domain protein in noncontact-inhibited cultured cells and postmitotic neurons. *Genes Dev.* **1**, 482-496.

Odenwald, W. F., Garbern, J., Arnheiter, H., Tournier-Lasserve, E. and Lazzarini, R. A. (1989). The Hox-1.3 homeo box protein is a sequence-specific DNA-binding phosphoprotein. *Genes Dev.* **3**, 158-172.

Orchard, K. and May, G. E. (1993). An EMSA-based method for determining the molecular weight of a protein-DNA complex. *Nucleic Acids Res.* **21**, 3335-3336.

Paro, R. (1990). Imprinting a determined state into the chromatin of *Drosophila*. *Trends Genet.* **6**, 416-421.

Pearce, J. J. H., Singh, P. B. and Gaunt, S. J. (1992). The mouse has a Polycomb-like chromobox gene. *Development* **114**, 921-929.

Pelaz, S., Urquia, N. and Morata, G. (1993). Normal and ectopic domains of the homeotic *Sex combs reduced* of *Drosophila*. *Development* **117**, 917-923.

Popperl, H. and Featherstone, M. S. (1993). Identification of a retinoic acid response element upstream of the murine *Hox-4.2* gene. *Molec. Cell. Biol.* **13**, 257-265.

Popperl, H., Bienz, M., Studer, M., Chan, S.-K., Aparicio, S., Brenner, S., Mann, R. S., and Krumlauf, R. (1995). Segmental expression of *Hoxb-1* is controlled by a highly conserved autoregulatory loop dependent upon *exd/pbx*. *Cell* **81**, 1031-1042.

Puschel, A. W., Balling, R. and Gruss, P. (1991). Separate elements cause lineage restriction and specify boundaries of *Hox-1.1* expression. *Development* **112**, 279-287.

Qian, S., Capovilla, M. and Pirrotta, V. (1991). The *bx* region enhancer, a distant cis-control element of the *Drosophila Ubx* gene and its regulation by *hunchback* and other segmentation genes. *EMBO J* **10**, 1415-1425.

Randazzo, F. M., Khavari, P., Crabtree, G., Tamkun, J. and Rossant, J. (1994). *brg1*: a putative murine homologue of the *Drosophila brahma* gene, a homeotic regulator. *Dev. Biol.* **161**, 229-242.

Riley, P., Carroll, S. and Scott, M. (1987). The expression and regulation of *Sex combs reduced* protein in *Drosophila* embryos. *Genes Dev.* **1**, 716-730.

Roy, R. J., Gosselin, P. and Guerin, S. J. (1991). A short protocol for micro-purification of nuclear proteins from whole animal tissue. *BioTechniques* **11**, 770-777.

Scott, M. P. and Weiner, A. J. (1984). Structural relationships among genes that control development: sequence homology between the *Antennapedia*, *Ultrabithorax*, and *fushi tarazu* loci of *Drosophila*. *Proc. Natl. Acad. Sci.* **81**, 4115-4119.

Sham, M. H., Vesque, C., Nonchev, S., Marshall, H., Frain, M., Gupta, R. D., Whiting, J., Wilkinson, D., Charnay, P. and Krumlauf, R. (1993). The zinc finger gene *Krox20* regulates *HoxB2* (*Hox2.8*) during hindbrain segmentation. *Cell* **72**, 183-196.

Shashikant, C. S., Bieberich, C. J., Belting, H.-G., Wang, J. C. H., Borbely, M. A., Ruddle, F. H. (1995). Regulation of *Hoxc-8* during mouse embryonic development: identification and characterization of critical elements involved in early neural tube expression. *Development* **121**, 4339-4347.

Struhl, G. (1989). Differing strategies for organizing anterior and posterior body pattern in *Drosophila* embryos. *Nature* **338**, 741-744.

Tuggle, C. K., Zakany, J., Cianetti, L., Peschle, C. and Nguyen-Huu, M. C. (1990). Region-specific enhancers near two mammalian homeo box genes define adjacent rostrocaudal domains in the central nervous system. *Genes Dev.* **4**, 180-189.

van der Lugt, N. M. T., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., van der Valk, M., Deschamps, J., Sofroniew, M., van Lohuizen, M. and Berns, A. (1994). Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the *bmi-1* proto-oncogene. *Genes Dev.* **8**, 757-769.

Vieille-Grosjean, I., and Huber, P. (1995). Transcription factor GATA-1 regulates human *HoxB-2* gene expression in erythroid cells. *J. Biol. Chem.* **270**:4544-4550.

Vogels, R., Charite, J., de Graaff, W. and Deschamps, J. (1993). Proximal cis-acting elements cooperate to set *Hoxb-7* (*Hox-2.3*) expression boundaries in transgenic mice. *Development* **118**, 71-82.

White, R. A. and Lehmann, R. (1986). A gap gene, *hunchback*, regulates the spatial expression of *Ultrabithorax*. *Cell* **47**, 311-321.

Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P. W. J., Stott, D. and Allemann, R. K. (1991). Multiple spatially specific enhancers are required to reconstruct the pattern of *Hox-2.6* gene expression. *Genes Dev.* **5**, 2048-2059.

Wolpert, L. (1989). Positional information revisited. *Development* (Suppl.), 3-12.

Zakany, J., Tuggle, C. K., Patel, M. D. and Nguyen-Huu, M. C. (1988). Spatial regulation of homeobox gene fusions in the embryonic central nervous system of transgenic mice. *Neuron* **1**, 679-691.

Zhao, J. J., Lazzarini, R. A. and Pick, L. (1993). The mouse *Hox-1.3* gene is functionally equivalent to the *Drosophila Sex combs reduced* gene. *Genes Dev.* **7**, 343-354.

## APPENDIX

### Construction of *Hoxa-5*/LacZ Mutant DNA for Generation of Transgenic Mice

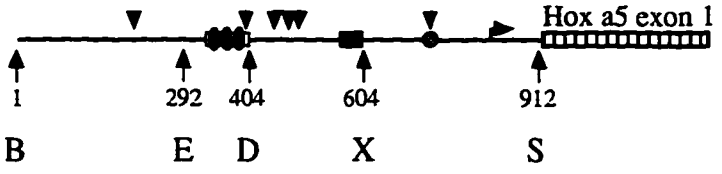
In order to generate the *Hoxa-5*/LacZ construct containing a mutation in site B only, primer 2, described in materials and methods, was used in a PCR reaction with an upstream primer U: 5'AGATCTTCCAGGCTGGATAAATAAACAAA containing an intact BglII site on the 5' end. This generated a 408bp product covering approximately two/thirds of the 5' region of the 604bp enhancer, containing a B mutation. After isolating the 408bp Bmut product is was used as a primer along with a downstream primer D: 5'CTCGAGCGCCACCCGCTG containing an intact XhoI site on the 3' end to generate the entire 604bp enhancer containing the B mutation with intact BglII and XhoI sites. The 604bp product was isolated and ligated into the pT7Blue vector from Novagen, p2-1 and sequenced.

To generate the same construct with all three sites mutated, A, B and C, an 83bp product was generated by PCR using primers 3 and 4 described in material and methods. This product was isolated and used as a megaprimer in a subsequent PCR reaction along with primer D to generate a 275bp product covering approximately the 3' one/third of the enhancer containing mutations is all three sites. Simultaneously primer 8: 5'GTATGGGGTCTgggcTTTGTGTAGAGG, containing a mutant A site, was used in a PCR reaction with primer U to generate the 350bp product covering approximately the 5' half of the enhancer. Both the 3' 275bp and the 5' 350bp products were ligated into the pT7Blue vector, p275-3 and p350-2, respectively, and sequenced. These vectors were then used as template in additional PCR reactions to re-amplify these regions. Primers U and 8 amplified the 5' 350bp region and primers D and 4 amplified the 3' 275bp region which contain overlapping sequence. The isolated products were used along with primers U and D in PCR in order to generate the entire 604bp region containing A, B and C mutant sites which was also ligated into the pT7Blue vector, p2-401, and sequenced.

The Bmut and ABCmut enhancers were then digested out of p2-1 and p2-401, respectively, by BglII and XhoI and isolated. We plan to ligate these fragments into pBal45Δ211 (see Figure 1, construct 2). Unfortunately the restriction sites in this vector do not allow a straight forward ligation and use of linkers to the isolated fragments has proven to be unsuccessful. Therefore, we have ligated the ABCmut fragment into pET17b (Novagen). Ligation into this vector was performed via BamHI and XhoI and this fragment can now be digested out with KpnI and XhoI sites. After isolation this fragment can be ligated into a dephosphorylated pBal45 vector digested with KpnI and XhoI. Using the NotI sites inpBal45 a construct containing the enhancer, with mutant A, B and C sites, the *Hoxa-5* promoter, LacZ and the SV40 polyadenylation site can be digested out and prepared for injection to generate transgenic mice. The LacZ expression will then be analyzed for an altered pattern in comparison to the LacZ expression pattern of transgenics carrying the same construct with no mutations (Figure 1 and 2, construct 2).

Figure 1. Deletion and addition analysis in transgenic mice defines cis-acting *Hoxa-5* regulatory elements. (A) Schematic map of the *Hoxa-5* 5' flanking region including the BSC enhancer (BSCE) previously identified (BglII to XhoI). Numbering begins at the BglII site (1). (B) *Homeobox*-lacZ gene fusions microinjected into fertilized oocytes to generate transgenic mice. At the left, various *Hoxa-5* BSCE regions used in the context of the *Hoxa-5* proximal promoter fused to the lacZ gene are shown. In constructs 1-9, exon-intron structure and poly A site was provided by SV40 sequences (Zakany et al., 1988), except for construct 5, which used *Hoxa-5* genomic region 2.8 Kb SacI-HindIII fragment (Tuggle et al., 1990). Some transgene constructs (bottom) also included the 2,800 bp *HOXD4* upper cervical enhancer. At the right is listed the number of expressing transgenic embryos dissected at e12.5, as well as the total number of transgenic embryos obtained for each construct. Arrowheads show Hox protein binding motifs (TAAT); white box and black ovals show the 333-404 region and AAATAA sites, respectively. Black box indicates the UCR element, and the shaded circle indicates the *Hoxa-5* binding site (Odenwald et al., 1989).* previously described data (Zakany et al., 1988); # expression in 2 of 18 was outside the spinal cord in the cervical and brachial regions of the peripheral nervous system. B: BglII; E: EcoRV; D: DdeI; X: XhoI; S: SacI.

A Restriction Map of Hox a5 5' Flanking Region




B Constructs Used in Transgene Analyses

| Construct Number | Figure 2 Letter | a5 Promoter / lacZ / pA | BSC Expression | Total Transgenics |
|---|---|---|---|---|
| | | B E D X | 6 | 14* |
| 1 | | Δ185 | 7 | 19 |
| 2 | A,B | Δ211 | 4 | 22 |
| 3 | C | Δ237 | 0 | 15 |
| 4 | | Δ260 | 0 | 9 |
| 5 | | Δ292 | 0 | 10 |
| 6 | D,E | Δ50 | 3 | 23 |
| 7 | F,G | Δ100 | 0# | 18 |
| 8 | | Δ200 | 0 | 17 |
| 9 | | Δ312 | 0 | 15 |

All constructs below include: D4 Enhancer/a5 Promoter — lacZ — a5 genomic region

(variable insert)

| | | | BSC Expression | UCSC Expression | Total Transgenics |
|---|---|---|---|---|---|
| 10 | H | (1-604) | 2 | 2 | 3 |
| 11 | J,K | (1-404) | 6 | 6 | 7 |
| 12 | I | (1-292) | 0 | 3 | 4 |
| 13 | L,M | (292-604) | 4 | 4 | 7 |
| 14 | N,O | (333-404) | 4 | 9 | 13 |
| 15 | P,Q | (292-367) | 3 | 6 | 15 |

Figure 2. Transgenic embryos showing representative expression patterns observed for various constructs. See Figure 1 for details on transgenes used to generate above embryos and Methods for lacZ assays. A, B: construct 2; C: construct 3; D, E: construct 6; F, G: construct 7; H: construct 10; I: construct 12; J, K: construct 11; L, M: construct 13; N, O: construct 14; P, Q; construct 15.

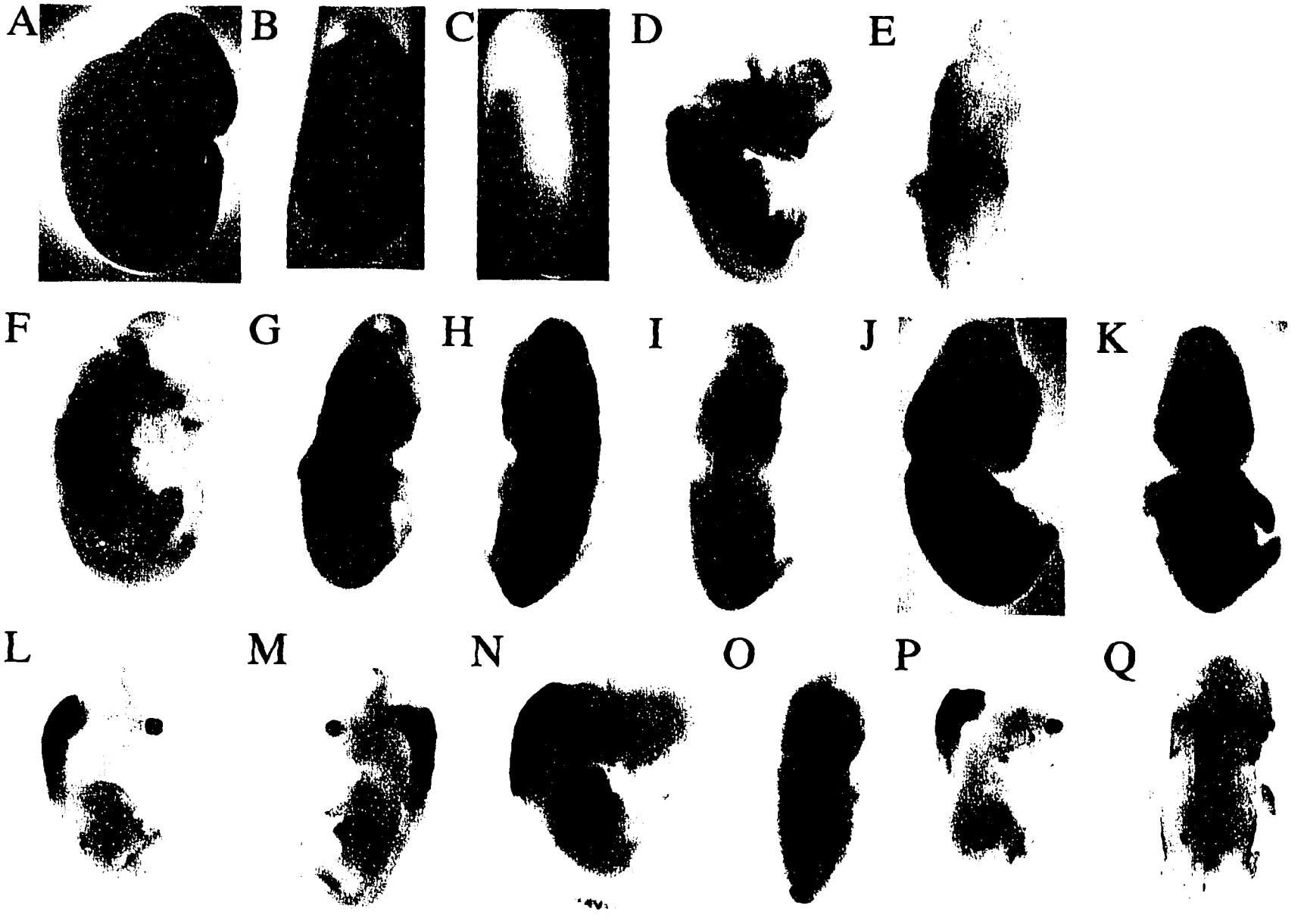
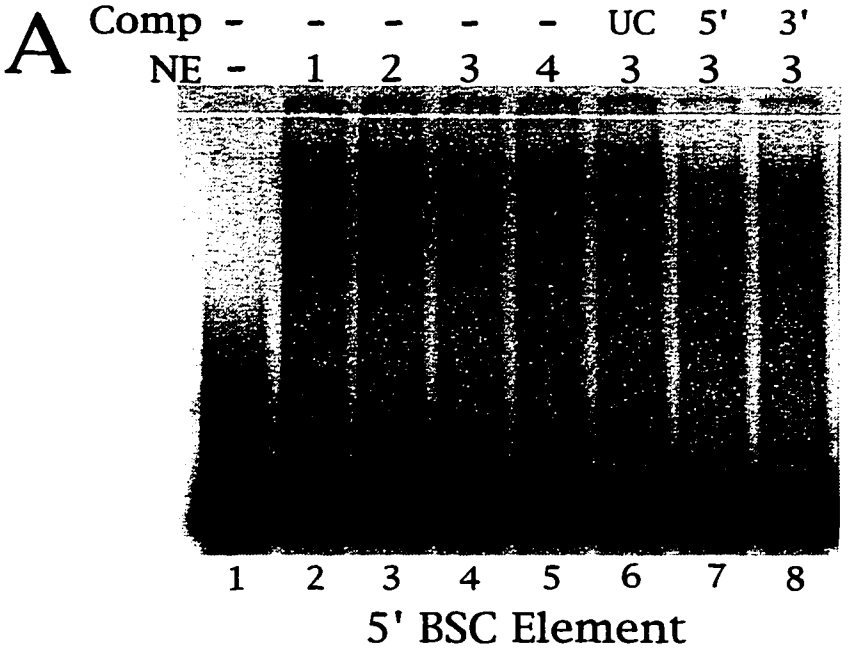

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q

Figure 3. EMSA analysis of the BSC element. Sequence-specific binding of e12.5 NE proteins with both the 5' (292-367) (A) and the 3' (333-404) (B) BSC regions. Lane 6 in each gel contains 200 molar excess oligo UCR element (UC) as nonspecific competitor and lanes 7 and 8 contain 200 molar excess 3' BSC and 5' BSC isolated fragment, respectively, as specific competitor. Lane 1- no extract, lanes 2, 3, 4, and 5- 2 mg of e12.5 NE of embryo sections 1, 2, 3, and 4, respectively. Specific activity of the 3' BSC probe was approximately 30% greater than that of the 5' BSC probe. Calculated molecular weight of the protein complexes binding in (B) are 89 and 122 kDa.


A
Comp - - - - - UC 5' 3'
NE - 1 2 3 4 3 3 3
1 2 3 4 5 6 7 8
5' BSC Element

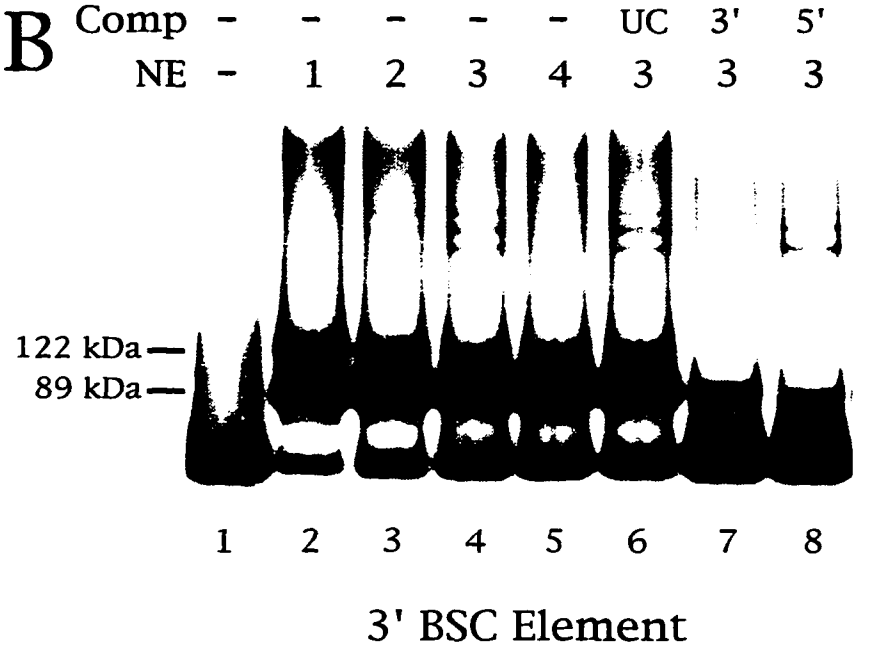
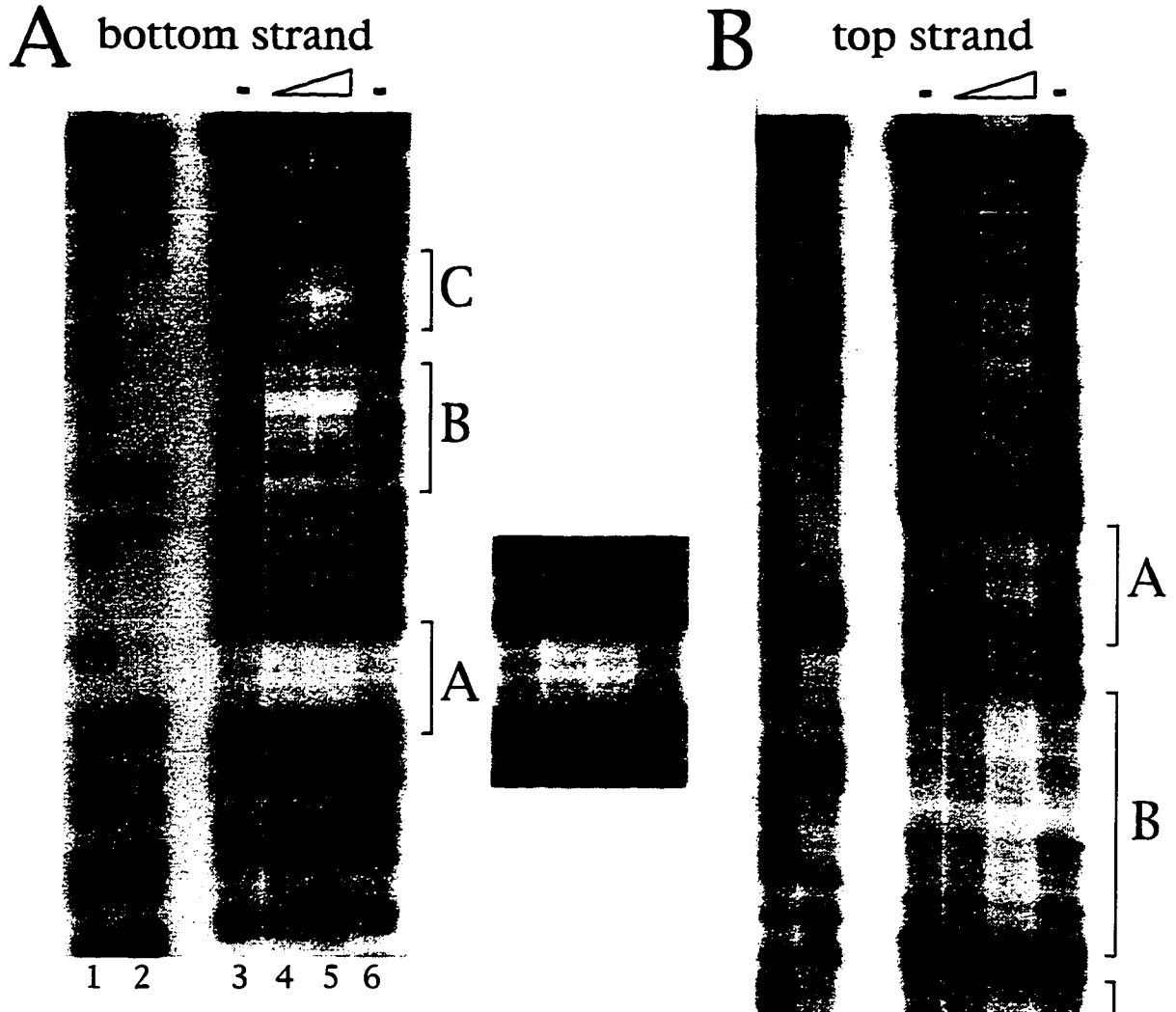

B
Comp - - - - - UC 3' 5'
NE - 1 2 3 4 3 3 3
122 kDa
89 kDa
1 2 3 4 5 6 7 8
3' BSC Element

Figure 4. DNaseI protection footprinting of the 3' BSC element. (A) Top strand. Lanes 1 and 2 contain A+G and G sequencing ladders, respectively. Lanes 3 and 6- no extract, lane 4- 400 ng fractionated whole e12.5 NE, lane 5- 2 mg fractionated whole e12.5 NE. Out-take to the right of top strand shows a darker exposure of the footprint site A. (B) Bottom strand. Lanes are the same as in (A). Brackets labeled A, B, and C designate observed footprints. (C) Protected sequence of footprints A, B, and C showing a 6 bp consensus of AAATAA. Brackets on top and below the sequence indicate the protected regions for the top and bottom strand, respectively. Letters in lower case above the sequence designate the bases mutated for the EMSA analysis in Figure 6.




C

Site A: ACAA**AAATAA**GAGG

Site B: GGG**AAATAA**AGTTG

Site C: GT**AAATAA**TTCTAAG

Consensus: GN**AAATAA**NNNTG

```
                                        gccc
5' GGGCTGGCAAACCGACCCCAACCTCTACACAAAAATAAGAGGGGATACAAA
   CCCGACCGTTTGGCTGGGGTTGGAGATGTGTTTTTATTCTCCCCTATGTTT
                                  Site A
           cccg      cgtc  cccg
   GCCGGGGAAATAAAGTTGTTGTAAATAATTCTAAGTCACCACCTCCCCCGA 3'
   CGGCCCCTTTATTTCAACAACATTTATTAAGATTCAGTGGTGGAGGGGGCT
         Site B           Site C
```

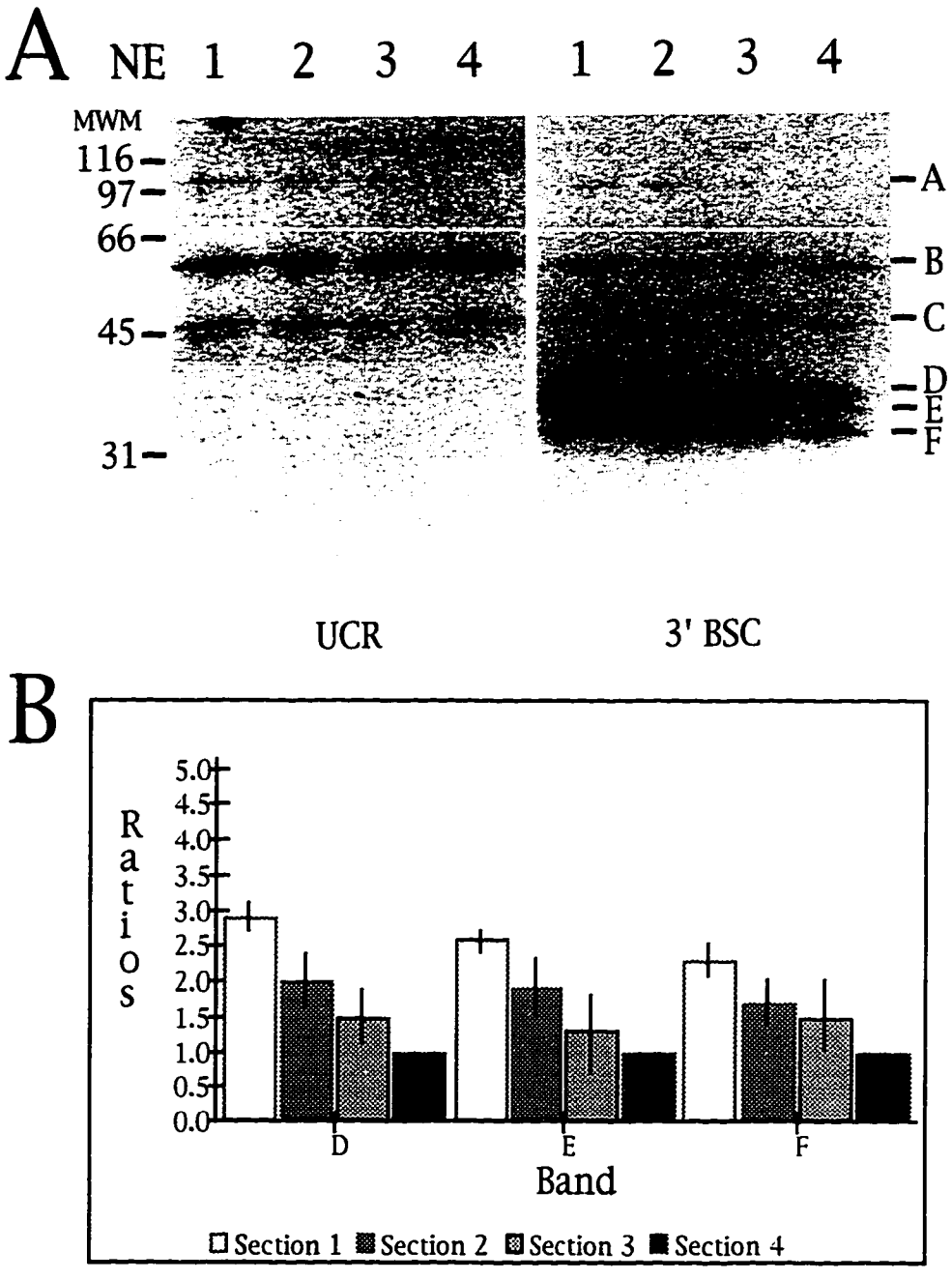
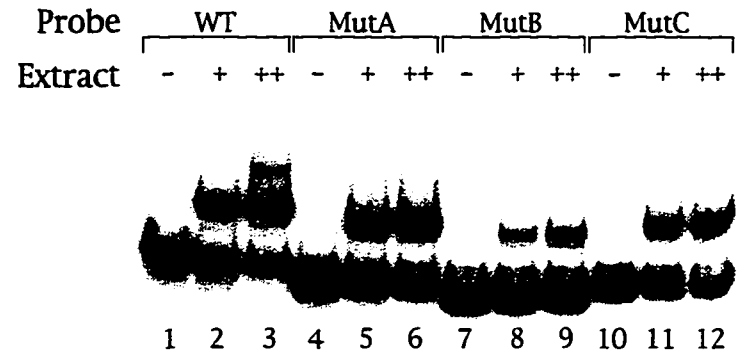


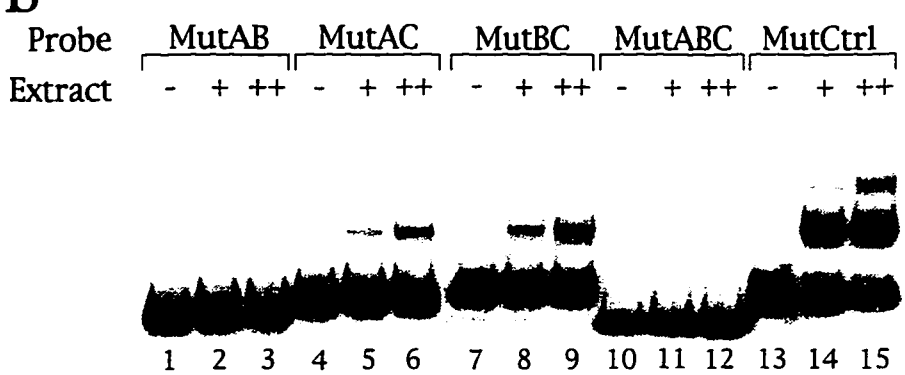

Figure 5. (A) Protein blot containing equal amounts (50 mg per lane) of e12.5 NE protein from each of the four sections probed with either the UCR element (left-hand blot) or the 3' BSC element (right-hand blot). Specific bands are labeled D, E, and F. (B) Each specific band was quantitated and normalized to a nonspecific band to assess differences in individual binding proteins to the BSC element across sections.

Figure 6. EMSA analysis using WT, MutA, MutB, MutC (A), MutAB, MutAC, MutBC, MutABC, and MutCtrl (B) probes. (A) Lanes 1, 4, 7, and 10- no extract; lanes 2, 5, 8, and 11- 100 ng of section 2 e12.5 NE; lanes 3, 6, 9, and 12- 200 ng of section 2 e12.5 NE. (B) Lanes 1, 4, 7,10 and 13- no extract; lanes 2, 5, 8,11 and 14- 100 ng of section 2 e12.5 NE; lanes 3, 6, 9,12, and 15- 200 ng section 2 e12.5 NE. (C) Percentage of bound DNA was quantitated for the WT and each of the mutants in (A) and (B). Percentage of activity relative to WT activity was calculated (see methods). Only the 200 ng extract lanes are shown on the graph, the 100 ng extract data showed a similar trend at approximately 50% of the values shown for the 200 ng extract lanes.

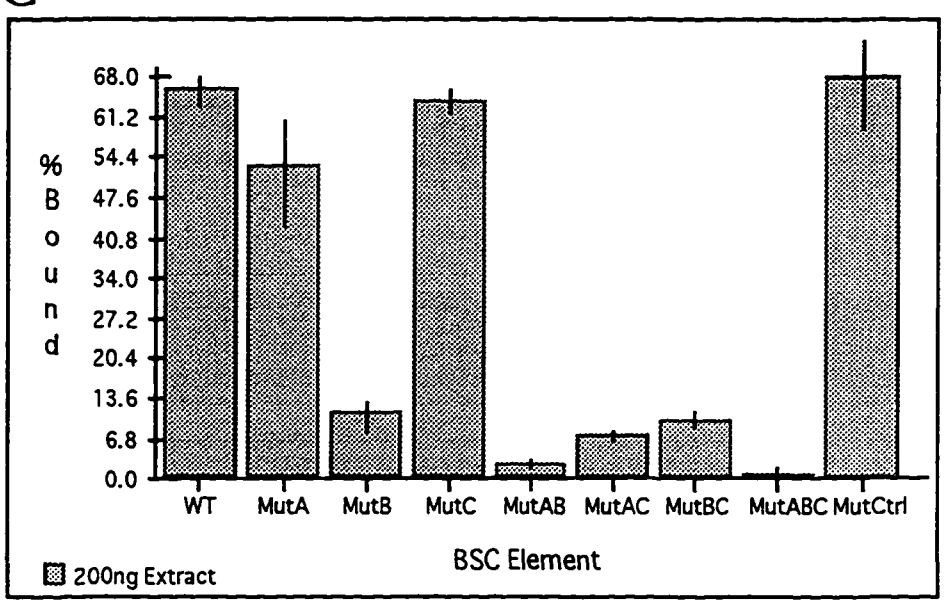
A

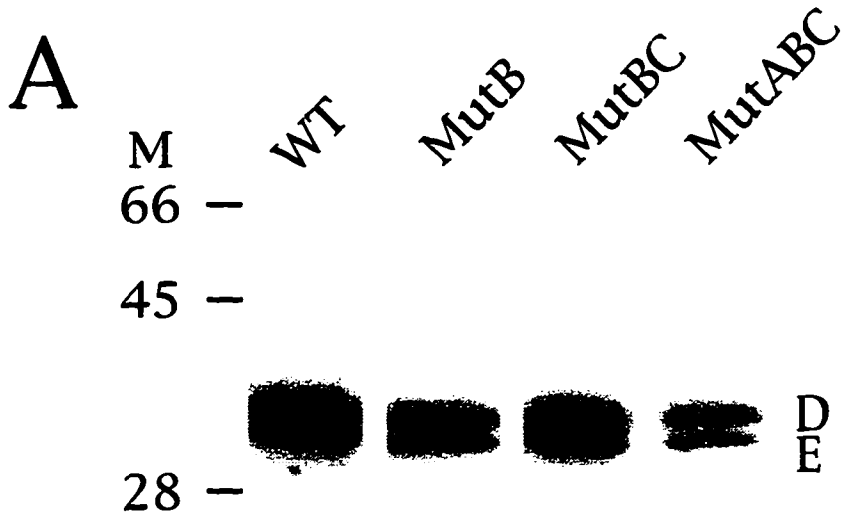

Probe
WT
MutA
MutB
MutC
Extract
- + ++ - + ++ - + ++ - + ++
1 2 3 4 5 6 7 8 9 10 11 12

B


Probe
MutAB
MutAC
MutBC
MutABC
MutCtrl
Extract
- + ++ - + ++ - + ++ - + ++ - + ++
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

C


% Bound
68.0
61.2
54.4
47.6
40.8
34.0
27.2
20.4
13.6
6.8
0.0
WT
MutA
MutB
MutC
MutAB
MutAC
MutBC
MutABC
MutCtrl
BSC Element
200ng Extract



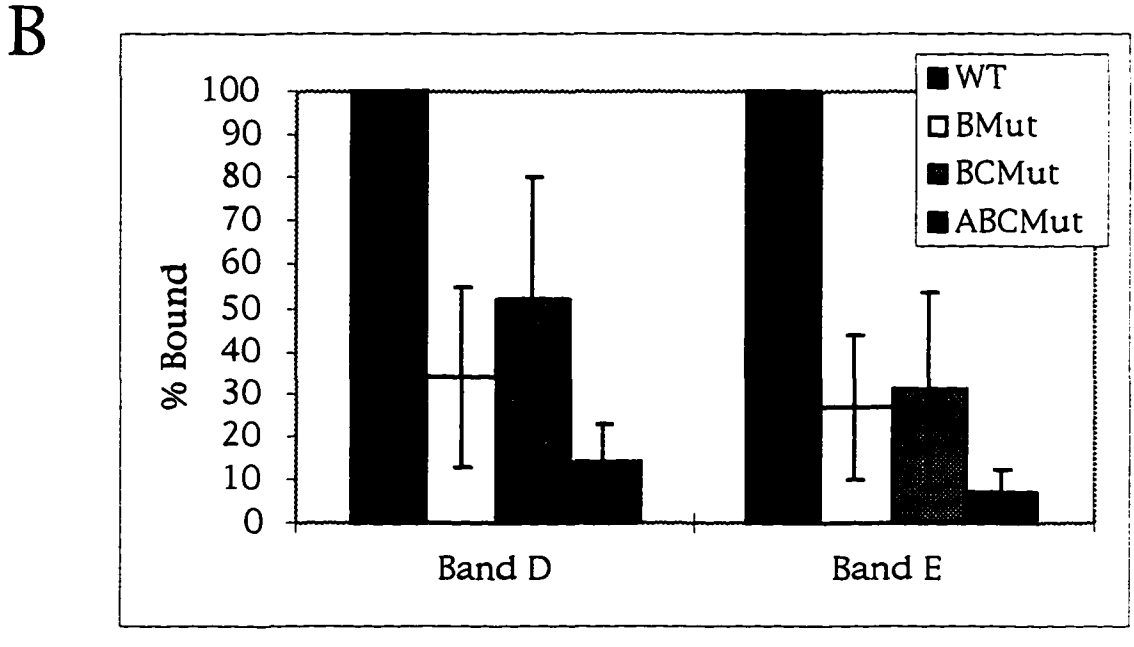




Figure 7. (A) Protein blot containing equal amounts (25 mg) of whole e12.5 NE. WT and Mutant probes labeled above each blot are the same as those used in the EMSA. (B) Binding activity of bands D and E with each probe were quantitated and calculated as a percentage of WT activity.

# CHARACTERIZATION OF PROTEINS BINDING TO TWO REGULATORY ELEMENTS OF *HOXA-5*

A paper to be submitted to the journal Mechanisms of Development

Tamara Nowling and Chris Tuggle

## ABSTRACT

Spatial regulation of the Hox genes has remain much of a mystery as only few direct regulators of these genes have been identified. Although regulation of the *Drosophila* homeotic genes, homologs of the Hox genes, is fairly well understood, extrapolation of this regulation to the mammalian genes has not been successful. In an attempt to understand Hox spatial regulation, we have been studying the activity of *Hoxa-5* reporter constructs in transgenic mice. LacZ driven by a 604bp enhancer of *Hoxa-5* expresses specifically in the brachial spinal cord region of 12.5 day mouse embryos. Previously we identified several cis-regulatory elements within this enhancer that are responsible for different aspects of this expression pattern and specific factors binding one of these elements, the brachial spinal cord element. We have now identified factors specifically binding another one of these elements, the upper cervical repressor element. We also show that DNA binding activity is temporally-specific for both of these elements as strong binding is observed from early gestation through midgestation but decreases thereafter and is absent just prior to birth. Additionally, we show that binding to the brachial spinal cord element is present in the same tissues in which *Hoxa-5* is specifically expressed and that this binding activity is not induced in the pheochromocytoma cell line PC12 by nerve growth factor. The temporal-specificity of binding and the fact that these factors are present in nuclear extracts from *Hoxa-5* expressing tissues are evidence that these factors may be involved in regulating *Hoxa-5*.

## INTRODUCTION

Spatial- and temporal-specific gene expression is central to the function of the homeotic and Hox genes in flies and mammals. These genes are found in clusters and are expressed colinearly in space and time which is dependent upon their position within the cluster. Hence, the more 3' genes in the cluster are generally expressed earlier with more anterior boundaries along the A-P axis of the embryo (Gaunt et al., 1988; Graham et al., 1989) and are induced earlier by retinoic acid (RA) and at lower concentrations of RA than the more 5' genes (Conlon et al., 1995). Many cis-acting regulatory elements have been identified in *Drosophila* homeotic genes (Kennison et al., 1993; Qian et al., 1991; Gindhart et al., 1995) and mammalian Hox genes (Zakany et al., 1988; Bieberich et al., 1990; Tuggle et al., 1990; Puschel et al., 1991;

Whiting et al., 1991; Gerard et al., 1993; Vogels et al., 1993; Brown and Taylor, 1994; Gutman et al., 1994; Shashikant et al., 1995; Charite et al., 1995) which are responsible for their expression patterns. And, although, trans-acting factors, which bind these cis regulatory elements of the homeotic genes in *Drosophila*, have been identified and regulation by these factors well characterized (Irish et al., 1989; White and Lehman, 1986; Kennison and Tamkun, 1992; Paro, 1990), very few factors have been identified in mammals and virtually nothing is known about the regulation of the Hox genes. Direct regulators of the Hox genes which have been identified include the retinoic acid receptors (RAR) (Popperl and Featherstone, 1993; Moroni et al., 1993; Langston and Gudas, 1992), Krox20 (Sham et al., 1993; Nonchev et al., 1996), pbx (Phelan et al., 1995; Chang et al., 1995; Chang et al., 1996), and the Hox genes themselves (Arconi et al., 1992; Popperl et al., 1995).

Functional retinoic acid response elements (RARE) have been identified near *Hoxa-1*, *b-1*, and *d-4* (reviewed in Conlon, 1995) In the NT2/D1 cell line most of the Hox genes in paralogous groups 1-8 are sequentially activated by RA, whereas those more 5' in the cluster (paralogous groups 9-13) are either downregulated or are undetectable after treatment with RA (reviewed in Boncinelli et al., 1991). It is believed that a concentration gradient of RA may exist in the CNS because the expression patterns of the Hox genes are also related to their position within the cluster (Maden and Holder, 1992). This inducibility may be a direct response to RA or a secondary response due to Hox regulators being induced by RA (Arconi et al., 1992). Enhancers found upstream of *Hoxb-2* (Sham et al., 1993) and *Hoxa-2* (Nonchev et al., 1996) contain binding sites for Krox20. These sites are required to drive LacZ expression to rhombomeres 3 (r3) and 5 (r5) in transgenic embryos. Pbx members interact cooperatively with several hox proteins in binding DNA in a specific manner; *HoxB-4*, *Hox A-1* and *HoxD-4* (Chang et al., 1995; Phelan et al., 1995). *HoxB-1* to *HoxB-9* will all interact with Pbx in binding a Hox half site but with decreasing preference from 3' to 5' within the cluster (Chang et al., 1996). As for the Hox genes themselves, hoxb-1 can autoregulate *Hoxb-1* but only in the presence of a coregulator, presumed to be Pbx (Popperl et al., 1995). Also, hoxC-6, D-10, D-9 and d-8 (more posteriorly expressed genes) but not hoxC-4 and D-4 (more anteriorly expressed genes) can transactivate *HoxC-5* (Arconi et al., 1992). This is consistent with the posterior prevalence or phenotypic suppression model which states that more posteriorly expressed genes are dominant (McGinnis and Krumlauf, 1992).

All of these Hox regulators, except Pbx, are themselves expressed in spatial- and temporal-specific manners. Pbx1, 2 and 3 mRNAs are expressed almost ubiquitously throughout the embryo. Transcripts are detected in most fetal and adult tissues and cell lines (Monica et al., 1991). Endogenous RA expression in the embryo is found in the ventral floor

plate of the neural tube (Wagner et al., 1990) and mRNA for the RARs is detected from embryonic day 7 to 9 in the neural tube beginning at the r3/r4 and r6/r7 boundaries for RARα and RARβ, respectively, and in the neural tube before the folds fuse for RARγ (Ruberte et al., 1991). Expression of the cellular retinoic acid binding protein (CRABP) is also detected in the neural tube with an r3/r4 boundary (Perez-Castro et al., 1989; Dencker et al., 1990; Ruberte et al., 1991; Maden et al., 1992). These boundaries coincide with Hox gene expression boundaries. Krox20 is specifically expressed in r3 and r5 on embryonic day 9.5 and is off in both rhombomeres by day 11 (Wilkinson et al., 1989) and, as mentioned earlier the Hox genes which have specific patterns of expression dependent upon the location within their cluster.

Cis-regulatory regions have been identified for spatial-specific expression of *Hoxa-5* to a subset of its endogenous pattern (Zakany et al., 1988; Tuggle et al., 1990; Nowling et al, unpublished). Because *Hoxa-5* has been extensively characterized in terms of its endogenous mRNA expression (Fibi et al., 1988; Zakany et al., 1988; Odenwald et al., 1987; Murphy et al., 1988; Dony and Gruss, 1987), protein expression (Tani et al., 1989), and function (Jeannotte et al., 1993; Aubin et al., unpublished), we are using this gene as a model for the spatial and temporal regulation of Hox genes in the CNS. Our previous work found specific proteins from embryonic nuclear extract (NE) binding to one of the identified regulatory elements, the brachial spinal cord (BSC) element in electrophoretic mobility shift (EMSA) and Southwestern assays (Nowling et al., unpublished). Here we report specific binding of proteins from embryonic NE to another regulatory element, the upper cervical repressor (UCR) element using similar assays. Using EMSA we show differential binding to these elements in NE from newborn and adult tissues and that the binding is temporally specific during gestation. Additionally, these binding proteins do not respond to NGF treatment in cultured cells.

## RESULTS

### Specific Binding of Embryonic Proteins to the *Hoxa-5* UCR Element

A 50bp upper cervical repressor (UCR) element was identified previously (Nowling et al., unpublished) which is necessary to prevent LacZ expression in the cervical region of the spinal cord in transgenic mice. In EMSA analyses with NE from whole e12.5 mice numerous shifted bands are detected with this element. Four of these shifts were determined to be specific in competition studies (Figure 1). Further analysis of two of these shifts, B and C, with NE from sections along the A-P axis of e12.5 mice detected slight differences in binding (Figure 2A). Binding in band B seems to be stronger in the anterior regions (sections 1-3) and much weaker in the posterior, while binding in band C is strongest in the cervical and brachial regions and weaker in the head and posterior regions (Figure 2B).

The proteins binding the UCR element were also analyzed throughout gestation from e9.5 to e18.5 for band B by EMSA. Results indicate that the complex in band B binds more strongly from e9.5-e14.5, but shows a drop in binding thereafter (data not shown). The degree of error was high due to the complexity of the banding pattern but using the two-tailed t-test there is a statistical difference in binding between midgestation (e13.5, 0.75% ± 0.26%) and late gestation (e18.5, 0.13% ± 0.12%) ($p < 0.005$).

**Analysis of the Complex Specifically Binding the *Hoxa-5* 3' BSC Element**

Specific proteins were previously identified to bind the *Hoxa-5* 3' BSC element (Nowling et al., unpublished). In order to characterize the temporal expression of the complex binding in the EMSA and the individual proteins binding on a protein blot, NE from different days of gestation was analyzed. By EMSA analyses, binding of the complex is observed from e9.5, the earliest day tested, to e14.5 and is absent by e18.5 (Figure 3A). This binding is strongest at e9.5 and again at e13.5 (Figure 3B). Binding of the individual proteins southwestern assays correlates with the EMSA in that binding is observed in e10.5-e14.5 NE but not in e18.5 NE (not shown, extract from e9.5 and e15.5-e17.5 were not tested), confirming the temporal binding specificity.

EMSA analyses was also utilized to observe tissue-specific expression of this complex. Figure 4 shows that binding is present in both newborn and adult kidney, liver, brain, lung, heart and spinal cord. In the newborn tissues the most intense binding is in the kidney followed by the liver, brain, lung and heart, although the binding in the liver is not seen as a sharp band as in the other tissues. Very little binding is observed in the spinal column. In the adult tissues intense binding is observed in the lung and the kidney with very little binding in the liver, brain, heart and spinal column (Figure 4).

**Binding Activity to the *Hoxa-5* BSC Element is Not Induced by NGF**

Because this complex binds to an element that is capable of expressing the LacZ gene in the spinal cord, one or more of the proteins in the complex may be induced by cells differentiating to a neuronal state, therefore increasing the level of binding of the complex to its response element. To test this, NE was made from PC12 cells treated with NGF and used in EMSA analyses. Binding is seen in either the uninduced or induced cell NE at 24 and 48 hours (Figure 5) with no change in the level of binding between uninduced and induced cells. Cellular response to NGF was confirmed at 48 hours by EMSA using a NGFRE probe and the same nuclear extract (data not shown). An approximate 4 fold increase of binding of specific nuclear proteins was observed in NE from cells induced for 48 hours compared to binding in NE from cells uninduced for 48 hours. An increase of binding activity of proteins from NGF-induced PC12 cell NE to this NGFRE has been previously reported (Higuchi et al., 1992).

## DISCUSSION

Endogenous *Hoxa-5* is spatially and temporally expressed in a specific manner during mouse development. Multiple transcripts are detected from the *Hoxa-5* locus (Fibi et al., 1988; Murphy et al., 1988; Zakany et al., 1988; Odenwald et al., 1987), but the lacZ expression pattern driven by the identified 604bp enhancer may represent the expression of the 1.8Kb proximal promoter because only the proximal promoter is present on the *Hoxa-5*/LacZ construct. These multiple transcripts are specifically expressed from e8-e13 during gestation (Dony and Gruss, 1987; Zakany et al., 1988, Fibi et al., 1988) in the spinal column, ribs and vertebrae, lung, kidney, stomach, and gut (Dony and Gruss, 1987; Fibi et al., 1988; Zakany et al., 1988; Odenwald et al., 1987). In addition, transcripts have been detected in newborn and adult tissues as well, including lung, liver, kidney, spinal column, and brain (Zakany et al., 1988; Odenwald et al., 1987). Based on the temporal- and tissue-specific expression of known regulators of *Hox*; RAR, Krox-20, and other Hox genes, we might expect the proteins binding to the cis regulatory elements of *Hoxa-5* to also be expressed in a temporal- and tissue-specific manner. Our prior results (Nowling et al., unpublished) and the results presented here provide some evidence for this hypothesis.

With both the UCR and BSC elements, the specific binding activity is present from e9.5-e14.5 in EMSA analyses. Between days e14.5 and e16.5 binding drops off and, for the BSC element binding, is absent by e18.5. This observation is confirmed for the BSC element by southwestern analyses which shows binding in e10.5-e14.5 NE but not in e18.5 NE. Southwestern analyses was not informative with the UCR element as no specific binding was observed by this method (Nowling et al., unpublished). This temporal-specific binding of proteins to these two regulatory elements corresponds with the temporal expression of *Hoxa-5* mRNA as detected by *in situ* hybridization (e8-e13) (Dony and Gruss, 1987) and by Northerns (e8-e14.5) (Fibi et al., 1988; Zakany et al., 1988), as well as the *Hoxa-5*/lacZ experiments (e11-e13) (Tuggle et al., 1990).

Binding of proteins to the BSC element in NE from newborn and adult tissues corresponds with the expression pattern of *Hoxa-5*, although these proteins seem to be more widespread. *Hoxa-5* mRNA (all transcripts) are present in newborn lung, kidney, and spinal column (Zakany et al., 1988) and specific binding to the BSC element is observed in these tissues also. But, we also detected binding in NE from liver, brain and heart. Duplicate preparations of NE showed similar ratios of binding activity (3-4 fold lower than shown in Figure 4) between newborn tissues, except for lung (~2 fold lower). In the adult, the *Hoxa-5* proximal transcript was detected in liver, kidney, spinal cord and brain (Odenwald et al., 1987). The binding activity varied between extract preparations of the adult tissues which may

be due to using different strains of mice for harvesting the adult tissues. Regardless of this variance, specific binding to the BSC element was observed in the same tissues that express *Hoxa-5* mRNA. In addition, binding was observed in NE from lung and heart.

Unpublished results of Aubin et al. show that the *Hoxa-5-/-* mice produced by Jeannotte et al. (1993) die due to laryngotracheal defects and lung immaturity. Our results show that the most intense binding to the BSC element is in NE from the adult lung. Strong binding was also observed in NE from the newborn lung. Additionally, these *Hoxa-5-/-* mice had altered expression of TTF-1, HNF3α and HNF3β which regulate surfactant proteins (Aubin et al., unpublished). Therefore, these proteins binding the BSC element may be involved in a pathway which regulates HNF3α/β and surfactant proteins of the lungs by directly regulating *Hoxa-5*. We expected that the most intense binding to the BSC element would occur in spinal column NE because the BSC element can confer brachial spinal cord expression to another enhancer/promoter LacZ construct in midgestational embryos (Tuggle et al., 1990; Nowling et al., unpublished). However, very little binding to the BSC element was observed in the spinal column extracts from both the newborn and adult

Based on the lacZ expression pattern in the spinal cord of transgenic mice which is driven by the *Hoxa-5* enhancer (Tuggle et al., 1990; Nowling et al., unpublished), and in an attempt to further characterize the proteins binding the BSC element of the enhancer, we tested whether this binding activity is induced in response to cells differentiating to a neuronal state by NGF. Induction did not occur in the NGF-induced cells as binding was observed in NE from uninduced as well as from induced with similar activity levels. Several conclusions can be drawn from these results. First, in agreement with the tissue extract data, these proteins are not likely to be neural specific. Secondly, because the binding activity is present in uninduced cell NE these may not sufficient for expression of *Hoxa-5*. Our interpretation is that there are other regulatory factors necessary for expressing *Hoxa-5* in its spatial-specific pattern based on some of our previous results which show that this complex binds the BSC element in NE from the head (Nowling et al., unpublished) while *Hoxa-5* is not expressed in the head. Alternatively, these proteins could be involved in maintaining or refining of expression rather than activation of expression because the binding activity of the embryonic proteins is highest on e13.5.

In conclusion, the temporal-specific binding of proteins during gestation to both the UCR and BSC cis-regulatory elements of *Hoxa-5* correlates well with the temporal-specific expression of *Hoxa-5*. The tissue specificity of proteins binding the BSC element in the newborn and adult is more widespread than seen for *Hoxa-5*. These results are consistent with our previous results (Nowling et al., unpublished) where we observed protein binding to the BSC element in all regions of the embryo, including the head and upper cervical region, while

transgenic mice only express β-galactosidase in the brachial region. However, these proteins are detected in all the tissues where *Hoxa-5* is expressed and binding activity is strong in the lung where lethal defects occur due to lack of *Hoxa-5*. Also, a recent paper by Mizuta et al. (1996) shows that rat *Hoxa-5* is expressed in normal and regenerating livers after hepatectomy and we find BSC binding proteins in NE from newborn and adult liver. Hence, the activity of the proteins binding the BSC element are consistent with the involvement of *Hoxa-5* in lung and liver development. These results are consistent with our hypothesis stated previously that *Hoxa-5* spatial-specific expression is regulated not only by the proteins we see binding to the BSC element but by other factors necessary for either repressing expression in the anterior regions of the embryo or additional proteins to allow expression in the brachial region, or both (Nowling et al., unpublished). Possible proteins that fit the profile include other, more anteriorly-expressed Hox genes that would be dominant to *Hoxa-5* expression and therefore, repress *Hoxa-5* expression in anterior regions where their expression domains overlap. The identification of a negatively acting element in the enhancer (Nowling et al., unpublished) is consistent with this idea. These proteins may be involved in the direct spatial-specific regulation of *Hoxa-5*, either in the initial activation or in the maintenance of *Hoxa-5* expression. Based on our results showing rather widespread expression of these proteins, regulation of other Hox genes by these proteins may also occur. Identification of the individual proteins in these binding complexes could help in understanding the spatial- and temporal-specific regulation of Hox genes. One possibility is *Cdx1* as its binding motif has some homology to the identified AAATAA binding sites on the BSC element. Several Hox gene regulatory regions contain putative *Cdx1* binding motifs and lack of *Cdx1* results in altered expression domains of Hox genes (Subramanian et al., 1995).

## MATERIALS AND METHODS

### Electrophoretic Mobility Shift Assay

Crude nuclear extracts (NE) (Roy et al., 1991) were prepared from whole embryonic 9.5-18.5 day postconception (e9.5-e18.5) mice, various newborn and adult tissues, and from approximately $1x10^8$ cells of uninduced and NGF-induced PC12 cells. Additionally, NE from regions (1-4) along the A-P axis of e12.5 mice were prepared. Regions 1-4 include: 1- head, to the lower jaw; 2- upper cervical, from lower jaw to just anterior of the forelimbs; 3- brachial, from anterior of forelimbs to just posterior of the forelimbs; 4- rest of the torso. Probes used for the brachial spinal cord (BSC) EMSA were isolated as described previously (Nowling et al., unpublished). NGFRE, and oct probes were oligos synthesized by the DNA Sequencing and Synthesis Facility at Iowa State University and were annealed prior to use. Upper cervical repressor (UCR) element probes were either isolated from pUC5, UCR element ligated into the

XhoI site of pBluescriptII (Stratagene), or annealed oligos synthesized by the DNA Sequencing and Synthesis Facility at ISU. All probes were labeled either by the Klenow reaction or by the T4 Kinase reaction.

EMSA conditions for binding to the BSC and the NGFRE probes were as described previously (Nowling et al., unpublished and Higuchi et al., 1992. Binding reactions and EMSA for the UCR element and octamer motif were performed as described in Roy et al., 1991 with some modifications. Reactions included polyd(I:C), approximately 0.5ng probe, 2μg NE, incubated for 20 minutes at room temperature and run on a 5%, 29:1 acrylamide:bisacrylamide, 0.25X TBE nondenaturing gel. For competition studies, competitor DNA was added to the reactions prior to addition of the probe.

Phosphorimaging was used for quantitation of specific shifted bands. For the BSC and NGFRE probes the percentage of DNA bound for each probe was calculated by dividing the amount of radioactivity in the shifted band by the total amount of radioactivity in the lane. In the UCR EMSA the specific shifted bands were normalized to the oct1 shifting band as oct1 has been shown to be ubiquitously expressed. Quantitation results are based on the average of at least three trials.

**Protein Blot Analysis**

Southwestern analyses were performed as described previously (Nowling et al., unpublished) using 25 μg NE from e10.5-e14.5 and e18.5 mice as described above. Either a PCR-generated wild type 3' BSC element or an isolated 3' BSC fragment as described in Nowling et al. (unpublished) were used as probes.

**Induction of Cell Lines**

The pheochromocytoma cell line PC12 was treated with 5ng/ml nerve growth factor (NGF) for 24 and 48 hours (kindly provided by Dr. Marit Nilsen-Hamilton). Induction of the PC12 cells at 24 and 48 hours was indicated by a morphology change to a neuronal state and confirmed at 48 hours by binding studies of NE from the cells to an NGFRE binding site

**ACKNOWLEDGMENTS**

The authors thank Dianna Wirzbach for assistance with EMSA gels, Lee Bendickson for NGF-induction of PC12 cells, and Lucie Jeannotte for RA-induction of P19 cells.

**REFERENCES**

Arconi, L., Simeone, A., Guazzi, S., Zappavign, , Ninelli, E. and Mavilio, F. (1992). The upstream region of the human homeobox gene *HOX3D* is a target for regulation by retinoic acid and HOX homeoproteins. EMBO J 11, 265-277.

Bieberich, C.J., Utset, M.F., Awgulewitsch, A. and Ruddle, F.H. (1990). Evidence for positive and negative regulation of the *Hox-3.1* gene. Proc. Natl. Acad. Sci. USA 87, 8462-8466.

Boncinelli, E., Simeone, A., Acampora, D. and Mavilio, F. (1991). Hox gene activation by retinoic acid. Trends Genet 7, 329-334.

Brown, W. M. and Taylor, G. R. (1994). The 5'-sequence of the murine *Hox-B3* (*Hox-2.7*) gene and its intron contain multiple transcription-regulatory elements. Int. J. Biochem.. 26, 1403-1409.

Chang, C-P., Shen, W-F., Rozenfeld, S., Lawrence, H.J., Largman, C. and Cleary, M.L. (1995). Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. Genes and Dev. 9, 663-674.

Chang, C-P., Brocchieri, Shen, W-F., Largman, C. and Cleary, M.L. (1996). Pbx modulation of hox homeodomain amino-terminal arms establishes different DNA-binding specificities across the Hox locus. Molec. Cell Biol. 16, 1734-1745.

Charite, J., de Graaff, W., Vogels, R., Meijlink, F., and Deschamps, J. (1995). Regulation of the *Hoxb-8* gene: Synergism between multimerized cis-acting elements increases responsiveness to positional information. Dev. Biol. 171, 294-305.

Conlon, R.A. (1995). Retinoic acid and pattern formation in vertebrates. TIG 11, 314-319.

Dencker, L., Annerwall, E., Busch, C. and Eriksson, U. (1990). Localization of specific retinoid-binding sites and expression of cellular retinoic-acid-binding protein (CRABP) in the early mouse embryo. Development 110, 343-352.

Dony, C. and Gruss, P. (1987). Specific expression of the *Hox 1.3* homeo box gene in murine embryonic structures originating from or induced by the mesoderm. EMBO J. 6, 2965-2975.

Fibi, M., Zink, B., Kessel, M., Colberg-Poley, A.M., Labeit, S., Lehrach, H. and Gruss, P. (1988). Coding sequence and expression of the homeobox gene *Hox 1.3*. Development 102, 349-359.

Gaunt, S.J., Sharpe, P.T. and Duboule, D. (1988). Spatially restricted domains of homeo-gene transcripts in mouse embryos: relation to a segmented body plan. Development 104 Suppl, 169-179.

Gerard, M., Duboule, D. and Zakany, J. (1993). Structure and activity of regulatory elements involved in the activation of the *Hoxd-11* gene during late gastrulation. EMBO J 12, 3539-3550.

Gindhart, Jr., J. G., King, A. N. and Kaufman, T. C. (1995). Characterization of the cis-regulatory region of the *Drosophila* homeotic gene *Sex combs reduced*. Genetics 139, 781-795.

Graham, A., Papalopulu, N. and Krumlauf, R. (1989). The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. Cell 57, 367-378.

Gutman, A., Gilthorpe, J., and Rigby, P. W. J. (1994). Multiple positive and negative regulatory elements in the promoter of the mouse homeobox gene *Hoxb-4*. Mol. Cell. Biol. 14, 8143-8154.

Higuchi, H., Nakano, K. and Miki, N. (1992). Identification of NGF-response element in the rat neuropeptide Y gene and induction of the binding proteins. Biochem and Biophys Res. Comm. 189, 1553-1560.

Irish, V. F., Martinez-Arias, A. and Akam, M. (1989). Spatial regulation of the *Antennapedia* and *Ultrabithorax* homeotic genes during *Drosophila* early development. EMBO J 8, 1527-1537.

Jeannotte, L., Lemieux, M., Charron, J., Poirier, F. and Robertson, E.J. (1993). Specification of axial identity in the mouse: role of the *Hoxa-5* (*Hox1.3*) gene. Genes and Dev. 7, 2085-2096.

Kennison, J. A. and Tamkun, J. W. (1992). Trans-regulation of homeotic genes in *Drosophila*. The New Biologist 4, 91-96.

Kennison, J. A. (1993). Transcriptional activation of *Drosophila* homeotic genes from distant regulatory elements. Trends Genet. 9, 75-79.

Langston, A. W. and Gudas, L. J. (1992). Identification of a retinoic acid responsive enhancer 3' of the murine homeobox gene *Hox-1.6*. Mech. Dev. 38, 217-228.

Maden, M and Holder, N. (1992). Retinoic acid and development of the central nervous system. BioEssays 14, 431-438.

Maden, M., Horton, C., Graham, A., Leonard, L., Pizzey, J., Eriksson, U. and Siegenthaler, G. (1992). Domains of retinoic acid-binding protein I (CRABP I) expression in the hindbrain and neural crest of the mouse embryo. Mech. Devel. 37, 13-23.

McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and axial patterning. Cell 68, 283-302.

Mizuta, I., Ogasawara, N., Yoshikawa, H. and Sakoyama, Y. (1996). Identification of homeobox genes expressed during the process of rat liver regeneration after partial hepatectomy. Biochemical Genetics 34, 1-15.

Monica, K., Galili, N., Nourse, J., Saltman, D. and Cleary, M.L. (1991). PBX2 and PBX3, new homeobox genes with extensive homology to the human proto-oncogene PBX1. Mole. Cell Biol. 11, 6149-6157.

Moroni, M. C., Vigano, M. A. and Mavilio, F. (1993). Regulation of the human *HoxD4* gene by retinoids. Mech. Dev. 44, 139-154.

Murphy, S.P., Garbern, J., Odenwald, W.F., Lazzarini, R.A. and Linney, E. (1988). Differential expression of the homeobox gene *Hox-1.3* in F9 embryonal carcinoma cells. Dev. Bio. 85, 5587-5591.

Nonchev, S., Vesque, C., Machonochie, M., Seitanidou, T., Ariza-McNaughton, L., Frain, M., Marshall, H., Sham, M.H., Krumlauf, R. and Charnay, P. (1996). Segmental expression of *Hoxa-2* in the hindbrain is directly regulated by Krox-20. Development 122, 543-554.

Odenwald, W.F., Taylor, C.F., Palmer-Hill, F.J., Friedrich, V., Jr., Tani, M. and Lazzarini, R.A. (1987). Expression of a homeo domain protein in noncontact-inhibited cultured cells and postmitotic neurons. Genes and Dev. 1, 482-496.

Paro, R. (1990). Imprinting a determined state into the chromatin of *Drosophila*. Trends Genet. 6, 416-421.

Perez-Castro, A., Toth-Rogler, L.E., Wei, L-N. and Nguyen-Huu, M.C. (1989). Spatial and temporal pattern of expression of the cellular retinoic acid-binding protein and the cellular retinol-binding protein during mouse embryogenesis. Proc. Natl. Acad. Sci. USA 86, 8813-8817.

Phelan, M.L., Rambaldi, I. and Featherstone, M. (1995). Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. Molec. Cell Biol. 15, 3989-3997.

Popperl, H. and Featherstone, M. S. (1993). Identification of a retinoic acid response element upstream of the murine *Hox-4.2* gene. Molec. Cell. Biol. 13, 257-265.

Popperl, H., Bienz, M., Studer, M., Chan, S.-K., Aparicio, S., Brenner, S., Mann, R. S., and Krumlauf, R. (1995). Segmental expression of *Hoxb-1* is controlled by a highly conserved autoregulatory loop dependent upon *exd/pbx*. Cell 81, 1031-1042.

Puschel, A.W., Balling, R. and Gruss, P. (1991). Separate elements cause lineage restriction and specify boundaries of *Hox-1.1* expression. Development 112, 279-287.

Qian, S., Capovilla, M. and Pirrotta, V. (1991). The *bx* region enhancer, a distant cis-control element of the *Drosophila Ubx* gene and its regulation by *hunchback* and other segmentation genes. EMBO J 10, 1415-1425.

Roy, R. J., Gosselin, P. and Guerin, S. J. (1991). A short protocol for micro-purification of nuclear proteins from whole animal tissue. BioTechniques 11, 770-777.

Ruberte, E., Dolle, P., Chambon, P. and Morriss-Kay, G. (1991). Retinoic acid receptors and cellular retinoid binding proteins II. Their differential pattern of transcription during early morphogenesis in mouse embryos. Development 111, 45-60.

Sham, M. H., Vesque, C., Nonchev, S., Marshall, H., Frain, M., Gupta, R. D., Whiting, J., Wilkinson, D., Charnay, P. and Krumlauf, R. (1993). The zinc finger gene *Krox20* regulates *HoxB2* (*Hox2.8*) during hindbrain segmentation. Cell 72, 183-196.

Shashikant, C. S., Bieberich, C. J., Belting, H.-G., Wang, J. C. H., Borbely, M. A., Ruddle, F. H. (1995). Regulation of *Hoxc-8* during mouse embryonic development: identification and characterization of critical elements involved in early neural tube expression. Development 121, 4339-4347.

Subramanian, V., Meyer, B.I. and Gruss, P. (1995). Disruption of the murine homeobox gene Cdx1 affects axial skeleton identities by altering the mesodermal expression domains of Hox genes. Cell 83, 641-653.

Tani, M., Odenwald, W.F., Lazzarini, R.A. and Friedrich, V.L., Jr. (1989). Progressive restriction in the distribution of the *Hox-1.3* homeodomain protein during embryogenesis. J. Neuro. Res. 24, 457-469.

Tuggle, C.K., Zakany, J., Cianetti, L., Peschle, C. and Nguyen-Huu, M.C. (1990). Region-specific enhancers near two mammalian homeo box genes define adjacent rostrocaudal domains in the central nervous system. Genes and Dev. 4, 180-189.

Vogels, R., Charite, J., deGraffe, W. and DesChamps, J. (1993). Proximal cis-acting elements cooperate to set *Hoxb-7* (*Hox-2.3*) expression boundaries in transgenic mice. Development 118, 71-82.

Wagner, M., Thaller, C., Jessell, T. and Eichele, G. (1990). Polarizing activity and retinoid synthesis in the floor plate of the neural tube. Nature 345, 819-822.

White, R. A. and Lehmann, R. (1986). A gap gene, *hunchback*, regulates the spatial expression of *Ultrabithorax*. Cell 47, 311-321.

Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P.W., Scott, D. and Alleman, R.K. (1991). Multiple spatially specific enhancers are required to reconstruct the pattern of *Hox 2.6* gene expression. Genes and Dev. 5, 2048-2059.

Wilkinson, D.G., Bhatt, S., Chavrier, P., Bravo, R. and Charnay, P. (1989). Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. Nature 337, 461-464.

Zakany, J., Tuggle, C.K., Patel, M.D. and Nguyen-Huu, C. (1988). Spatial regulation of homeobox gene fusions in the embryonic central nervous system of transgenic mice. Neuron 1, 679-691.

## APPENDIX

P19 cells were induced with $10x1^{-6}$ M retinoic acid (RA) for 1, 3, 5, 7, and 14 hours (kindly provided by Dr. Lucie Jeannotte). NE was prepared the same as for the PC12 cells discussed in Materials and Methods. Induction of P19 cells using this protocol has been shown by assaying for *Hoxa-5* mRNA which was induced between 3 and 6 hours after RA addition (L. Jeannotte, personal communication), but was not assayed for on this particular set of cells. Because *Hoxa-5* is indirectly induced by RA, the binding activity to the BSC element may be directly induced by RA. By using EMSA, the level of binding activity to the 3' BSC element was quantitated as a way to test for induction of this complex. Figure 6A shows binding of the complex in uninduced as well as in RA-induced P19 cell NE. Although no increase in binding activity was observed in NE from uninduced cells compared to cells induced for 2-14 hours (Figure 6B), we cannot draw any definite conclusions because we have not been able to obtain a second set of cells or positively confirm induction of this set of cells as we have no positive control as we did with the NGF-induced cell NE. An attempt to show induction by RA for this set of cells was performed by assaying for an increase in binding of hoxa-5 to its binding site located in the *Hoxa-5* promoter. Unfortunately, we were unable to identify which shift was due to binding of hoxa-5.

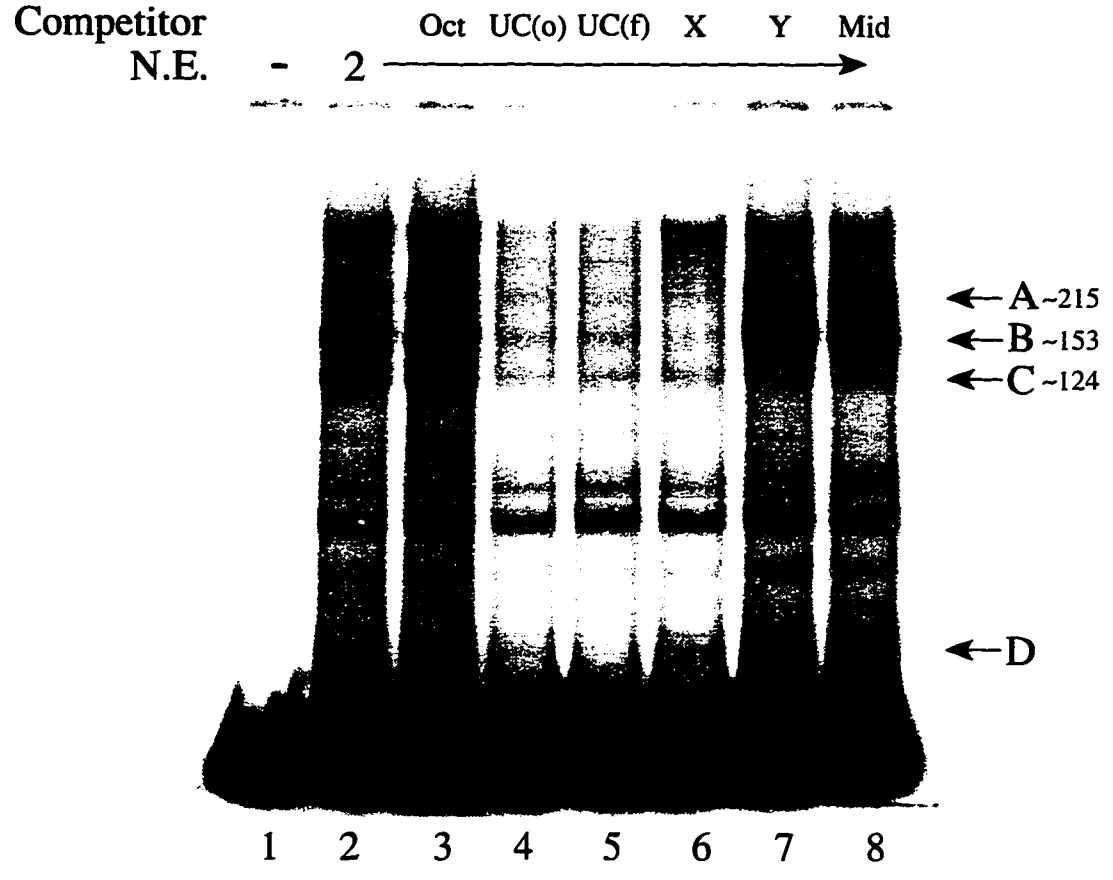
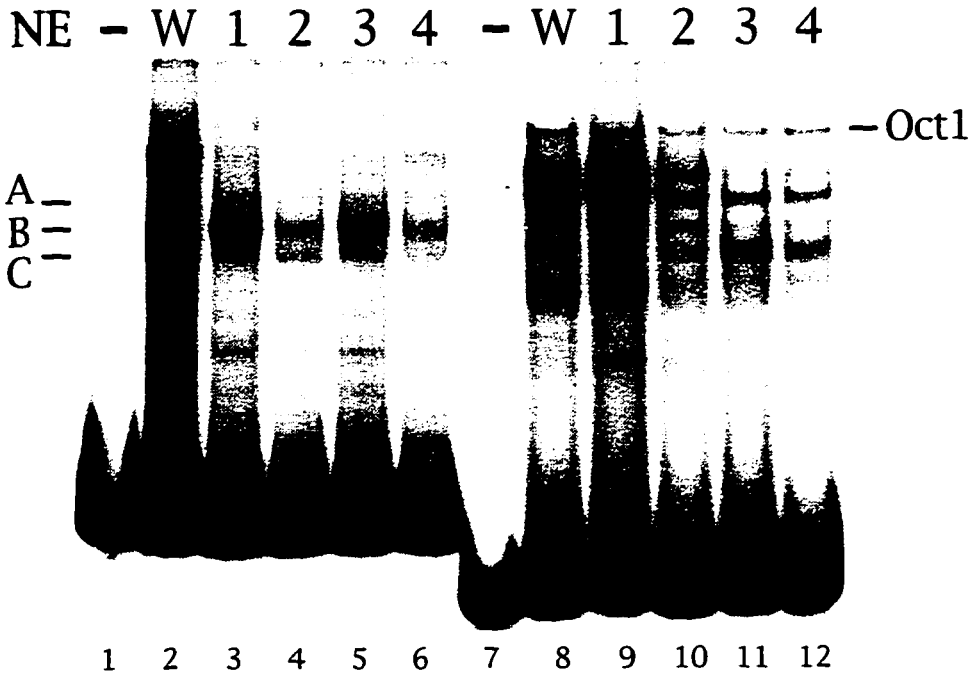



Figure 1. EMSA analysis of the UCR element. Sequence-specific binding of proteins in a NE from e12.5 mice. Lane 1 contains no extract and all subsequent lanes contain 2 mg of e12 NE of embryo section 2. The following lanes contain 200 molar excess of competitors: lane 3- oct oligo (oct), as nonspecific competitor; lane 4- oligo of the UCR element (UCo); lane 5- isolated fragment of the UCR element (UCf); lane 6- oligo of the left two-thirds of the UCR element (X); lane 7- oligo of the right two-thirds of the UCR element (Y); and lane 8- oligo of the middle third of the UCR element (Mid). Specific shifted bands are labeled A, B, C and D.

Figure 2. EMSA analysis of the UCR element. Differences in binding across the A-P axis of the 12.5 day embryo. (A) Lanes 1-6 run with an isolated fragment of the UCR element as probe and lanes 7-12 run with an oligo of the oct motif as probe. All subsequent lanes contain 2 mg of the following NE; lane 1 and 7- no extract, lanes 2 and 8- whole embryo (W), lanes 3 and 9- section 1 (1), lanes 4 and 10- section 2 (2), lanes 5 and 11- section 3 (3), and lanes 6 and 12- section 4 (4). (B) Specific bands B and C were quantitated for each section and normalized to the Oct1 band to assess differences in binding activity across the A-P axis of the embryo.

A


NE - W 1 2 3 4 - W 1 2 3 4
Oct1
A
B
C
1 2 3 4 5 6 7 8 9 10 11 12

B

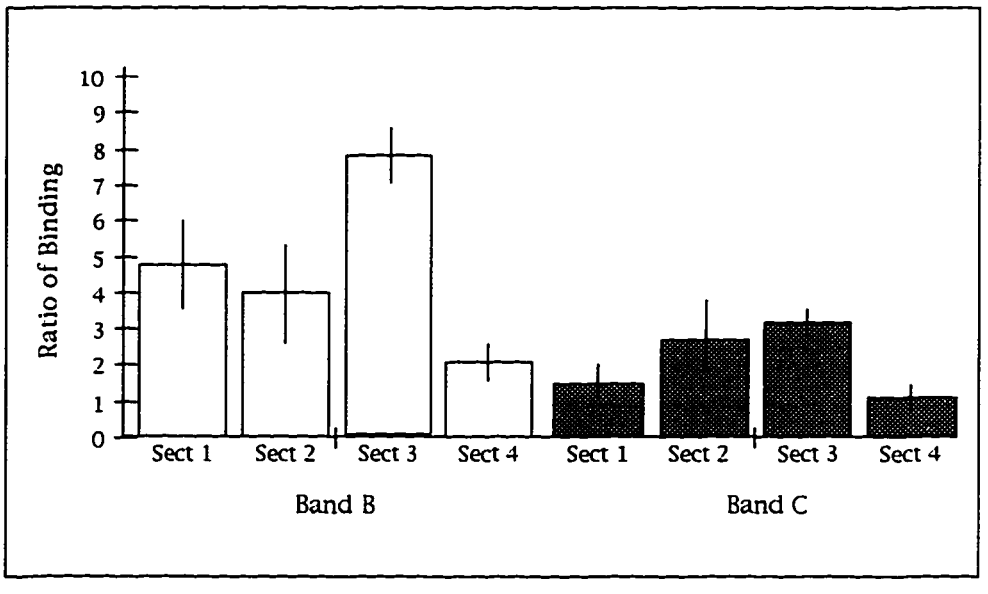
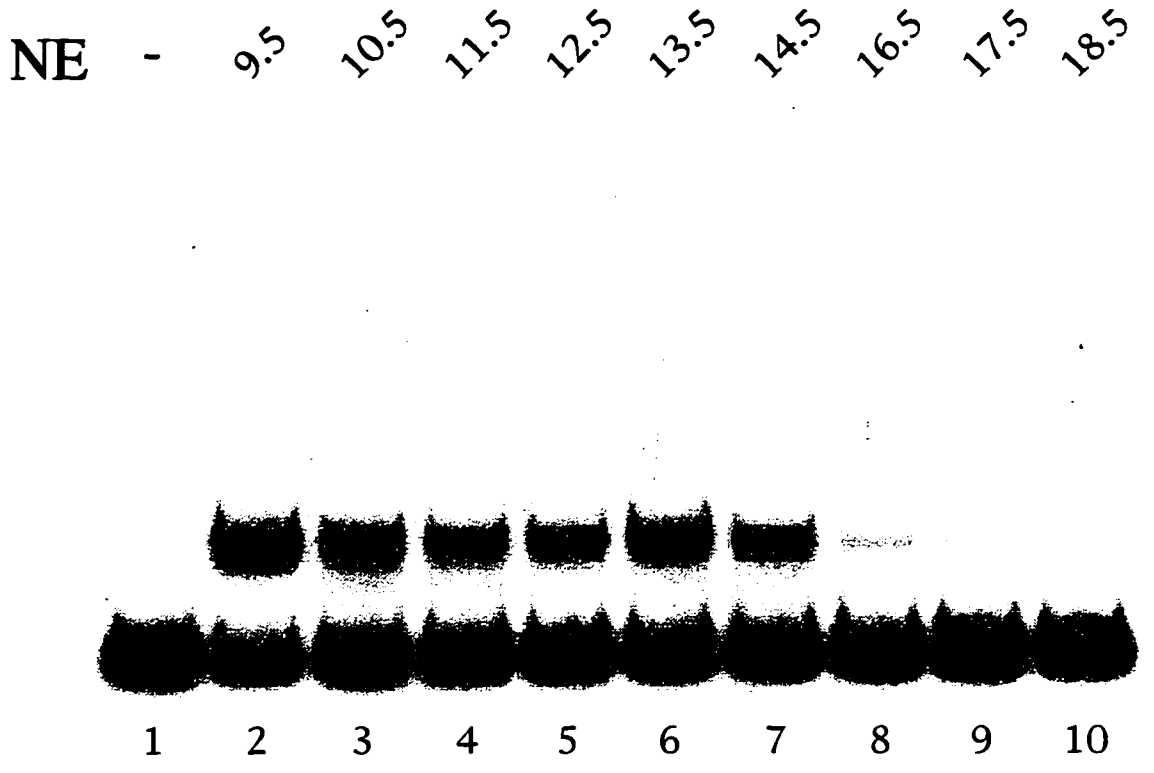
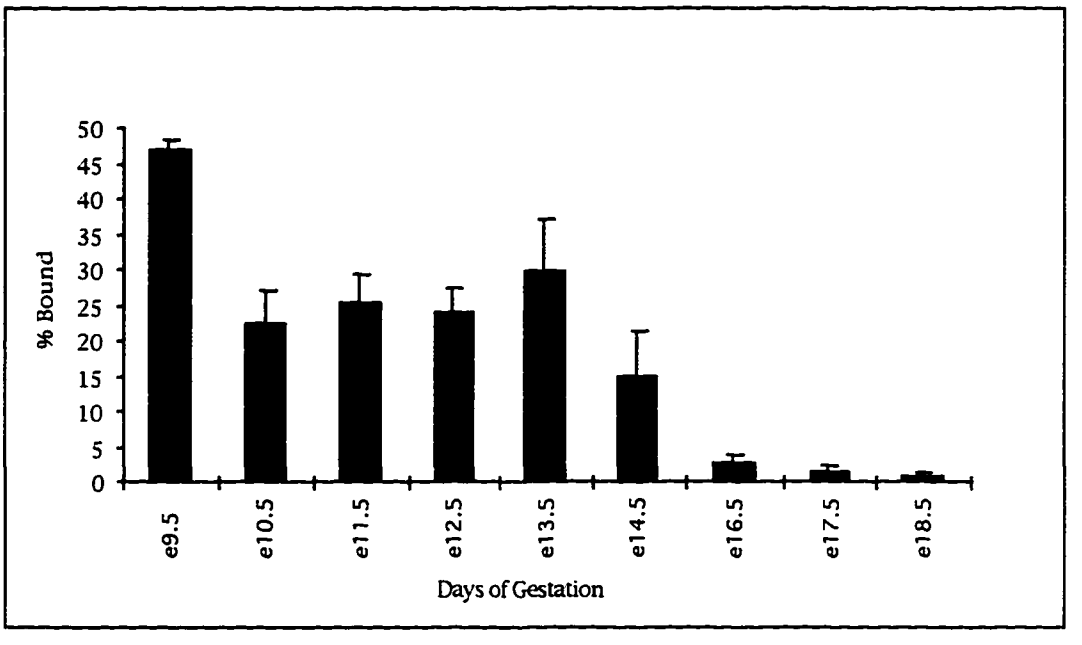

Ratio of Binding
10
9
8
7
6
5
4
3
2
1
0
Sect 1
Sect 2
Sect 3
Sect 4
Sect 1
Sect 2
Sect 3
Sect 4
Band B
Band C

Figure 3. EMSA analysis of specific binding to the BSC element during gestation. (A) Lane 1 contains no extract. Lanes 2-10 contain 50 ng of whole embryonic NE from embryos of the following age respectively; 9.5, 10.5, 11.5, 12.5, 13.5, 14.5, 16.5, 17.5, and 18.5. (B) Specific binding was quantitated for each day of gestation to assess temporal differences in binding activity. Calculations are based on three trials with one preparation of NE. EMSAs were also performed with a second preparation of NE which resulted in similar levels of binding activity.

A


NE
-
9.5
10.5
11.5
12.5
13.5
14.5
16.5
17.5
18.5
1
2
3
4
5
6
7
8
9
10

B


% Bound
50
45
40
35
30
25
20
15
10
5
0
e9.5
e10.5
e11.5
e12.5
e13.5
e14.5
e16.5
e17.5
e18.5
Days of Gestation

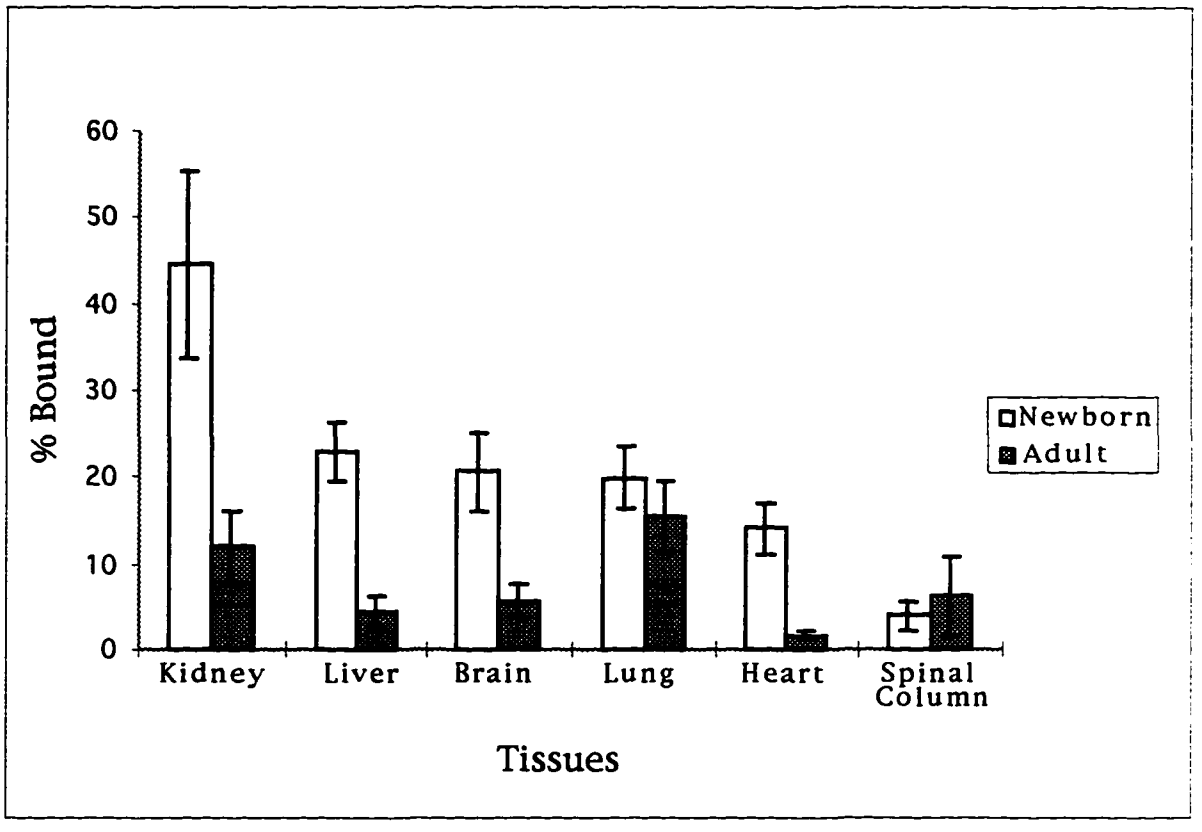



Figure 4. Quantitation from EMSA analyses of specific binding to the BSC element in various newborn and adult tissue NE. Tissues include lung, kidney, liver, heart, brain and spinal column. Specific binding was quantitated for each tissue to assess tissue specific differences in binding activity. Calculations are based on four trials from one preparation of NE. EMSAs were also performed on a second extract preparation which resulted in similar ratios of binding activity between newborn tissues, except for newborn lung. Binding in adult tissue NE varied between extracts.

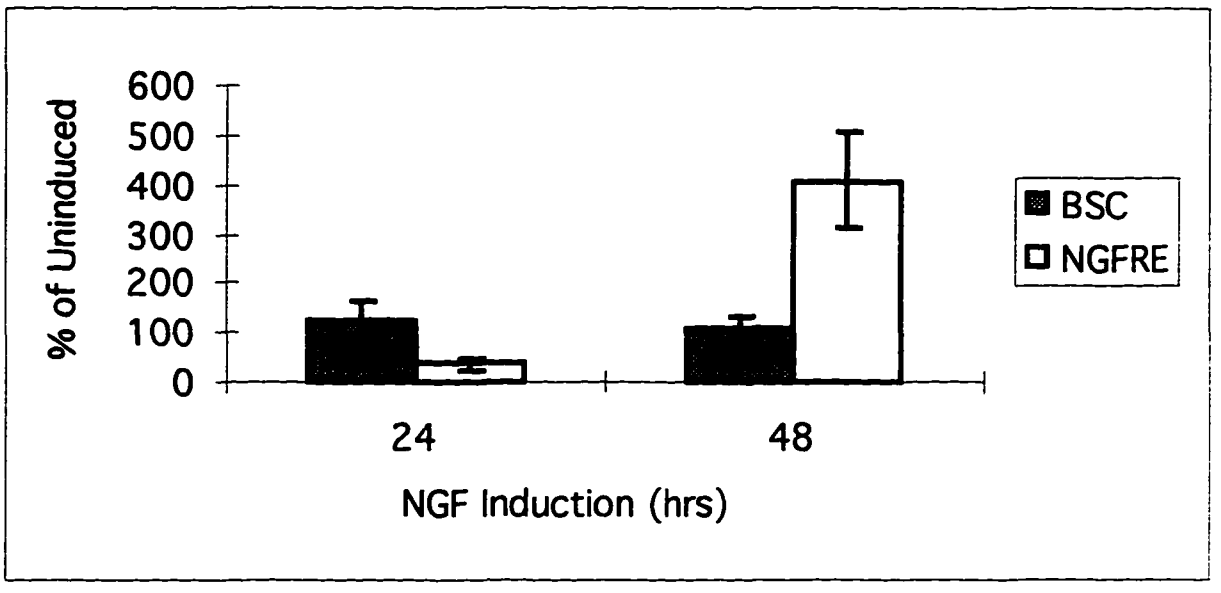



Figure 5. Quantitation from EMSA analysis of specific binding to the BSC and NGFRE elements in NGF-induced PC12 cell NE. PC12 cells were induced for 24 and 48 hours. Specific binding was quantitated and calsulated as a percentage of uninduced for each time point to assess for NGF-induction of binding activity. Only the 48 hour NGF treatment was found to have induced the cells as shown by the NGFRE binding activity. Calculations are based on three trials with one preparation of NE.

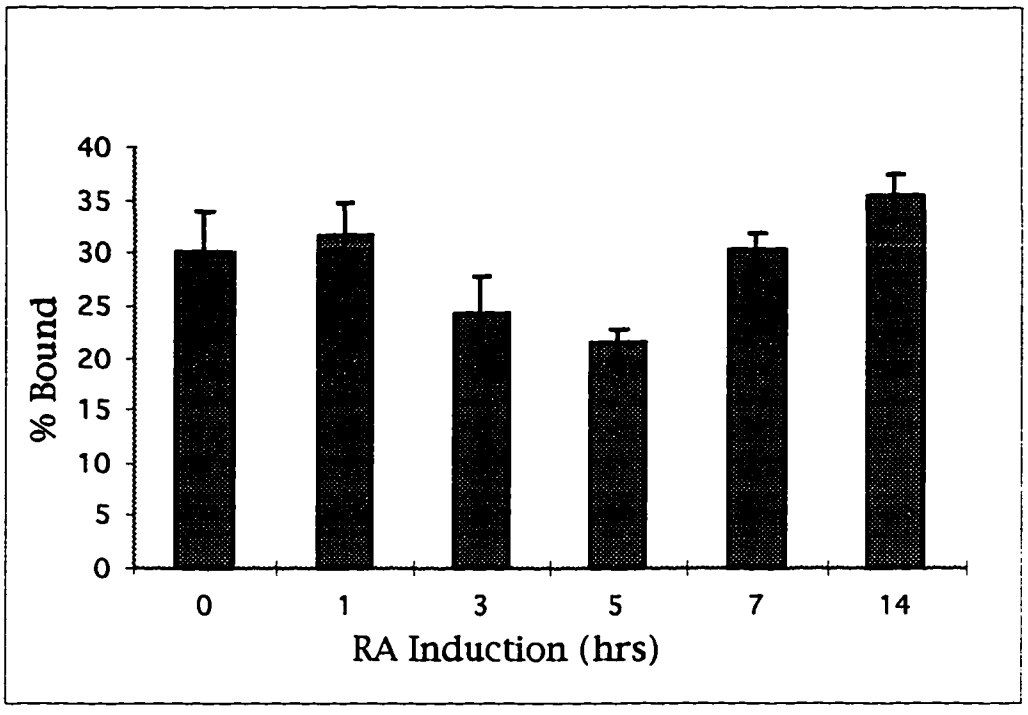



Figure 6. Quantitation from EMSA analysis of specific binding to the BSC element in RA-induced P19 cell NE. P19 cells were induced for 0, 1, 3, 5, 7 and 14 hours. Specific binding was quantitated for each time point to assess for RA-induction of binding activity. Calculations are based on four trials with one preparation of NE.

# GENERAL CONCLUSIONS

## Summary of Results

Regulatory elements, identified previously by Dr. Chris Tuggle and others (see first paper), have been found to be necessary for different aspects of the spatial-specific expression pattern of *Hoxa-5*/lacZ in transgenics. We hypothesized that transcription factors are specifically binding these elements and regulating the spatial-specific expression of *Hoxa-5* from the proximal promoter. In addition, these proteins may themselves be expressed in a temporal- and tissue-specific manner. In order to test this hypothesis we developed specific conditions to use several different binding assays for detection of embryonic proteins binding *Hoxa-5* regulatory elements. Electrophoretic mobility shift assay (EMSA) analyses showed specific binding to both the upper cervical repressor (UCR) element and the brachial spinal cord (BSC) element, whereas southwestern analyses showed specific binding of individual proteins to only the BSC element. Competition studies in the EMSA suggested that the specific binding to the UCR and BSC elements was contained within a 20bp region and a 25bp region, respectively. Binding in bands B and C (~153 and ~124 kDa, respectively) to the UCR element exhibits differences in activity levels across the A-P axis of the 12.5 day embryo. Binding activity in band A is higher in the anterior regions of the embryo and lowest in the posterior, whereas the binding activity in band B is somewhat lower in the head and tail and higher in the midsection of the embryo.

Specific binding in embryonic NE was also detected by southwestern analysis with the BSC probe. Individual proteins of between approximately 32 and 36 kDa were found to bind as a gradient from head (high) to tail (low). A survey of the expression of the proteins binding these elements was performed using EMSA and southwestern analyses. Temporally, strong binding activity to the UCR element is present from embryonic day 9.5-14.5 and declines until it is almost completely gone by day 18.5 on EMSAs. Similar temporal specificity is seen for the proteins binding the BSC element on EMSAs. Strong binding is present from embryonic day 9.5-14.5 with a peak around 12.5-13.5, very little binding is present by day 16.5 and binding is absent by day 18.5. The temporal binding specificity is reinforced by results from southwestern analyses which show that the individual proteins are observed to bind in NE from 10.5-14.5 day embryos but no binding is observed in NE from 18.5 day embryos. This temporal binding activity is consistent with the endogenous temporal expression of *Hoxa-5* which is expressed from e8-e14 (Fibi et al., 1988; Zakany et al., 1988), implying that these proteins are likely candidates in the regulation of *Hoxa-5*.

Using NE made from various newborn and adult tissues EMSAs were performed in order to observe the tissue-specific expression of the binding activity to the BSC element. Strong binding was observed in kidney, lung, liver and heart. In addition to the tissues, the inducibility of the binding activity by retinoic acid (RA) and nerve growth factor (NGF) was analyzed. NE was made from either RA-induced P19 cells or NGF-induced PC12 cells. Binding activity was present in uninduced cell NE and was not induced by either RA or NGF.

Now that specific binding to these elements has been identified, the sequences to which the proteins are binding need to be identified and mutations made within these sites in to characterize the necessity of these proteins in the regulation of *Hoxa-5*. DMS protection footprinting and DNaseI protection footprinting was performed. No protected sites were identified by DMS footprinting. However, three protected sites were found by DNaseI footprinting. All three of the sites contain a 6bp core sequence of AAATAA which does not have any significant homology to any other known DNA binding domain. The closest homology is with a Cdx binding domain. Because all three of these sites were A/T rich, protection in the DMS footprinting assay was not expected as only protection of G residues would be detected.

Mutations were made within this core AAATAA site by changing the central AATA to G/C residues by PCR. Subsequently, probes of the BSC element containing mutations in one, two or all three sites were used in EMSA and southwestern analyses. Results with both of these assays show that binding is reduced with single mutations to varying degrees depending on which site was mutated. Binding was reduced significantly more with the double mutations and completely abolished with the triple mutation in the EMSA and reduced by greater than 80% with the triple mutant in the southwestern.

These results indicate that the specific binding observed in both the EMSA and southwestern are due to the identified protected sequences containing the core sequence AAATAA. But, are these sites, and hence, the proteins binding them, important for the spatial-specific expression of *Hoxa-5 in vivo*? A construct containing the *Hoxa-5* enhancer with all three sites mutated, the *Hoxa-5* proximal promoter, the lacZ gene and the SV40 polyadenylation signal has been made and will be injected and transgenic mice analyzed. If this mutant construct results in the lack of LacZ expression or an altered LacZ expression pattern, then we can conclude that the proteins binding these sites are involved in the spatial-specific expression of *Hoxa-5*. This same construct, minus the first 211bp of sequence, containing entirely wild type sequence was used in transgenic analysis showing LacZ expression specifically in the brachial region of the spinal cord (see first paper).

## Discussion

The activity of the complex binding the BSC element during early to midgestation indicates that these proteins may be specifically regulating genes involved in patterning the embryo during midgestation, including the Hox genes. Although the binding activity of these proteins in various tissues indicates that they are more widespread than expected, our data indicates that the level of the individual factors within the binding complex varies in different regions and it may be these differences that contribute to the spatial regulation of *Hoxa-5*. Evidence for this was seen in the southwestern analyses where a gradient of binding activity is present along the A-P axis of the day 12.5 embryo. Because the strongest binding activity was observed in NE from the head and *Hoxa-5* is not expressed in the head, other factors must be involved in regulating the spatial expression of *Hoxa-5*. Additional evidence for the involvement of other factors is based on the EMSA results that indicate the complex binds in a slight gradient opposite that seen by individual proteins in the southwestern. A repressor protein may be present in the anterior regions of the embryo which binds another site on the enhancer (such as the UCR element) to prevent *Hoxa-5* expression in the head and upper cervical regions. Alternatively, another positive factor may exist in a gradient opposite to the factor(s) in the southwestern and both of these factors are necessary for *Hoxa-5* expression to begin in the brachial spinal cord. Both of these possibilities may also be true.

An alternative hypothesis is that these proteins may be involved in more of a maintenance role in regulating *Hoxa-5*, as well as other Hox genes. The binding activity of this complex in the EMSA indicates that these proteins are relatively abundant, as only 25ng of a crude embryonic NE is needed to see between a 20-40% shift of the free probe. This fact and the widespread nature of the binding activity in newborn and adult tissue NE could be an indication that these proteins are more general regulators. Additional evidence for a more general role is observed in the EMSA with NE from RA-induced P19 cells and NGF-induced PC12 cells. Strong binding activity is present in the uninduced and the induced cell NE. *Hoxa-5* is not induced in P19 cells until 3-6 hours of RA treatment (Lucie Jeannotte , personal communication), so why are these binding proteins present much earlier than *Hoxa-5* expression? If these proteins are involved in regulation/maintaining regulation of other genes, such as *Hoxa-1* which is induced very early by RA, then its presence in uninduced cell NE would be necessary to maintain that activation or perhaps to refine the spatial boundaries of expression after initial activation or possibly to help in the initiation of activation.

The binding activity detected in NE from newborn and adult lung is additional evidence that these proteins may be involved in the regulation of *Hoxa-5* based on data from *Hoxa-5* -/- mice. Homozygous *Hoxa-5* deficient mice die due to defects associated with lung physiology.

*Hoxa-5* -/- mice were found to have altered expression patterns of several regulators of surfactant proteins including TTF-1, HNF3α, and HNF3β (Aubin et al., unpublished), indicating that these proteins may be directly regulated by *Hoxa-5*. Binding proteins present in lung NE may therefore, be involved in the *Hoxa-5* regulatory pathway of the surfactant proteins by positively regulating/maintaining the expression of *Hoxa-5* in this tissue. It would be interesting to generate transgenic mice carrying a *Hoxa-5* gene with mutations for the identified binding sites in the BSC element and look at the expression of these surfactant protein regulators. If the expression of the surfactant proteins is altered we could conclude that proteins binding the BSC element are part of a pathway for regulating the expression of the surfactant proteins which involves *Hoxa-5*.

In an attempt to identify the individual proteins that are a part of the complex binding the BSC protected sites, we employed an expression library screen. The screen identified several clones with one open reading frame (ORF) in common, ORF2. We have shown that protein induced from ORF2 does bind the BSC element in southwestern assays and we have some evidence that it can bind in EMSA assays. Binding of ORF2 in the southwestern may be more robust due to the fact that the binding occurs on proteins immobilized on paper, which is how these clones were isolated, rather than in liquid because on paper there is a higher relative concentration of this protein(s) available for binding. A single protein that normally binds as part of a complex may not bind as strongly on its own and therefore, it may need to be present in a high concentration in order to force binding to its response element. Our current hypothesis is that ORF2 is a member of the complex from embryonic NE binding in the EMSA. This hypothesis can be tested by producing an antibody to ORF2 and using it in supershifting experiments. If these results are show ORF2 is present in the EMSA binding complex, then we have identified a new DNA binding protein and its response element as there is no homology of ORF2 to any known gene and no homology of the binding sites to any known response elements, other than Cdx. Additionally, ORF2 may be important for the spatial-specific expression of *Hoxa-5* from the proximal promoter, which can be confirmed or discounted based upon the results of the transgenics containing *Hoxa-5*/LacZ with mutations in the binding sites.

**Future Directions**

There are many experiments that can be done to further characterize the binding proteins and their binding sites on the BSC element in order to better understand the spatial-specific regulation of *Hoxa-5*. These include injecting the *Hoxa-5* mutant/LacZ construct which is in progress. Analysis of transgenic mice carrying this construct for β-gal expression will tell us if these sites are important for *Hoxa-5* regulation *in vivo*. If these sites prove to be important

then the same mutations could be "knocked in" to generate transgenic mice carrying an endogenous *Hoxa-5* deficient for these binding sites. Any resulting phenotype can be compared to the phenotypes seen in *Hoxa-5-/-* mice, especially the expression pattern of the surfactant proteins mentioned earlier which are altered in the *Hoxa-5 -/-* mice.

Cdx-/- embryos are being sent from Peter Gruss' lab in Europe. We can make NE from these embryos to test whether Cdx is in fact part of the binding complex since the Cdx binding site has homology to our identified binding sites. If this is the case then we should see little or no binding to the BSC element due to lack of Cdx in the NE. Additionally, antibodies can be made to ORF2 and used in EMSA analyses to test for supershifting due to the antibodies binding and increasing the molecular weight of the complex or for loss of binding due to the antibodies binding and interfering with ORF2 binding the response element. The antibodies can also be used in western analyses with NE from embryos during gestation and from different newborn and adult tissues to look at temporal- and tissue-specific expression of the ORF2 protein.

Northerns, in addition to or prior to westerns, can be performed using ORF2 sequence as probe on RNA from different days of gestation and from newborn and adult tissues to test temporal- and spatial-specific expression of ORF2 mRNA. These results can then be compared to the binding activity studies on NE from the same days of gestation and tissues, as well as to the western results. Along the same lines as the Northerns, *in situ* hybridizations would show the expression pattern of ORF2 within embryos. The protein itself could be characterized by making deletions of the ORF2 sequence in one of the bacterial expression vectors and protein extract made. Using the southwestern assay, these deletions of ORF2 can be used to identify the regions of the protein necessary for binding.

Because ORF2 is an unknown gene, regardless of its importance in regulating *Hoxa-5*, a genomic library screen or RT PCR of genomic mouse DNA can be performed in order to identify the full length gene/cDNA for ORF2. The gene could then be mapped or further characterized in a variety of ways. Characterization may include knocking out the gene and observing any phenotypes and confirming that this gene is involved in regulating *Hoxa-5* by looking for altered expression of *Hoxa-5*.

## References

Fibi, M., Zink, B., Kessel, M., Colberg-Poley, A.M., Labeit, S., Lehrach, H. and Gruss, P. (1988). Coding sequence and expression of the homeobox gene *Hox 1.3*. Development 102:349-359.

Zakany, J., Tuggle, C.K., Patel, M.D. and Nguyen-Huu, C. (1988). Spatial regulation of homeobox gene fusions in the embryonic central nervous system of transgenic mice. Neuron 1:679-691.

# APPENDIX I
# EXPRESSION LIBRARY SCREEN

Because individual proteins were able to bind to the BSC element in the southwestern analyses (see first paper), we decided to screen an expression library in the hopes of isolating and identifying one or more of the proteins binding the BSC element. An embryonic 11.5 day mouse λgt11 library from Clontech was screened with an isolated fragment of the 3' brachial spinal cord (BSC) element, end-labeled and ligated. Filters were soaked in 10mM IPTG to induce expression of protein fused to β-galactosidase (β-gal) from the cDNA clones. 600,000 phage were plated in the primary screen and 15 positives which showed binding on duplicates filters were selected. A secondary screen confirmed 9 of the positives. PCR was then performed on resuspended phage from agar plugs using the 3' and 5' gt11 arms as primers (Notebook V, pp. 131 and 135). Digests of these PCR products indicated that two of the positives were identical therefore, one was eliminated and the remaining 8 positives were used in a tertiary screen (Notebook V, pg.137). Single plaques were picked from the tertiary screen and PCR was repeated on the resuspended phage which confirmed the PCR results from the secondary screen (Notebook V, pg.147). Sizes were estimated as follows: A4- 4.5Kb, A2B- 2.6Kb, A3B- 2.1Kb, B2B- 1.2 and 1.6Kb (2 bands amplified), H1- 1.3Kb, H2- 2Kb, and K1C- 1.3Kb. The remaining clone, B1A failed to amplify and we eventually discontinued working with this clone.

Fusion protein extract was then made from all of these clones to test for binding in an EMSA. Protein was induced directly from plates of growing phage by adding 5ml of 10 mM IPTG, 10mM $MgSO_4$ in 0.5x Luria broth to the plates and incubating for 1 hour at 37°C (Huang et al., 1989; Huang and Jong, 1994). The liquid was then removed and dialyzed against and the supernatant was used in the binding reactions. This procedure was repeated six times and EMSAs performed on each extract preparation. We repeatedly saw a shifted band with the A2B extract different from any of the shifts seen with the other clone extracts that was able to be competed away with cold specific competitor but not with cold nonspecific competitors (UCR fragment and oligo of oct site) (Figure 1A and 1B). This shift was seen in 4 of the 6 extract preparations and was competed away by specific competitor in 3 of the 4 extracts. One EMSA was performed by Dr. C. Tuggle with the Bmut and ABmut probes as well as the wild type BSC probe (CKT Notebook )in order to check the specificity of the binding and to see if this protein is a part of the complex seen binding the A, B and C sites in embryonic NE (see first paper). The shift in the A2B extract lane was less intense with the ABmut than with the Bmut but a definite conclusion cannot be made about the specificity of the

binding because binding to the wild type (PCR generated) probe was basically nonexistent, although binding was observed with the same extract and an isolated fragment of the BSC element. With the help of Dr. C. Tuggle, lysogens of all the clones were made ( Notebook VI, pps. 36-49) following the protocol of Cowell and Hurst (pp. 120-122) and two preparations of fusion protein extract made from the lysogens following the protocol of Cowell and Hurst (pp. 120-122). EMSA analysis was then performed and the shifted bands observed were similar to those seen in the control lane (extract from Y1089 only). Protein gels run on the supernatant and pellet fractions of the extract preparations from the lysogens (Notebook VI, pg. 58) indicated that most of the protein was in the pellets so had precipitated out. In an attempt to renature the proteins from the pellets the second preparation of lysogen extract was resuspended in urea and then dialyzed to slowly remove the urea. The resulting supernatant was made 20% glycerol final and stored at -70°C. EMSA results on these denatured-renatured extracts were inconclusive as no specific binding was observed (Notebook VI, pg. 126).

Clone A2B, being one of the longer clones and showing specific binding in EMSAs, was cloned into the pBluescriptII vector. Claire Schmitz then helped to make EXOIII/S1 deletions of the clone (Notebook VII, pp. 3-12 and Notebook VI, pg. 120) so the entire A2B clone could be sequenced. In addition to these deletions, all clones were sent to the sequencing and synthesis facility at Iowa State University and sequenced in on the 3' and 5' ends using the same primers used for PCR. The entire sequence for A2B was obtained and sequence on the 3' and 5' ends for the rest of the clones (Notebook VII). Three open reading frames (ORF) are present in A2B. ORF1 is 171 amino acids (aa), ORF2 is 215 aa and ORF3 is 137 aa. The 5' ends of all of the clones except K1C and B2B could be found within the A2B clone (Figure 2). The smallest clone, H1 is contained entirely within A2B and contains all of ORF2 and part of ORF3. Additionally, we discovered that none of these clones were actually fused to β-gal. Hence, these proteins must have been translated due to reinitiation.

Initial dot blot analysis performed by C. Schmitz showed that all of the clones could hybridize to the A2B and A4 clones when they were used as probes. This assay was repeated a second time with the A2B clone with the same results (Notebook VI, pp.29 and 57). These probes did not hybridize to clones isolated from a phage pig library indicating that the probes were specifically hybridizing to the cDNA sequence and not the phage. In addition to the dot blot analysis, southern blots were performed with the help of C. Schmitz and Dr. C. Tuggle (CKT Notebook X, pg.52). The first Southern contained all of the clones and was probed with either a HindIII fragment, from the 5' end of A2B, or a DraI fragment, from the 3' end of A2B. Only A2B hybridized to the HindIII fragment and all of the clones except K1C and B1A hybridized to the DraI fragment. These results have some inconsistencies with the sequence

information. None of the other clones overlap with the 5' end of the A2B which is consistent with the sequence information, but all of the clones except B1A and K1C hybridize to the DraI fragment on the 3' end of A2B including H1 which ends within the A2B sequence so it does not contain the DraI fragment sequence (Figure 2). It is interesting to note that according to the sequencing results, K1C and B2B do not begin or end anywhere within A2B, yet K1C hybridizes to A2B and A4 by dot blot analysis and B2B also hybridizes to the DraI fragment on the Southern (Notebook VI, pp.29 and 57). Another Southern was run by Dr. C. Tuggle containing an array of digestions of mouse genomic DNA and probed with the entire H1 clone. H1 hybridized to one band in BamHI, EcoRI and PstI digestions, two bands in a HindIII digestion and three bands in the HaeII digestion (Figure 3), indicating this clone is most likely a single copy gene in the mouse. A zoo blot was performed by Diana Walker which showed binding to mouse DNA of about the same size as the previous southern in an EcoRI digest, but no binding to DNA from other organisms (DW Notebook, pp. 41-57). A blast search with the A2B protein sequence resulted in a some homology of ORF1 to a DNAB protein, a binding protein, and no homology with the other two open reading frames. Results with A2B DNA sequence showed no significant homology with anything in the databases (for blast results see Notebook VII).

The EMSA results indicated that extract from the A2B clone gave a specific shift but, several open reading frames are found within this clone and we needed to test which one is responsible for that shift. Because multiple clones were isolated all containing ORF2, we proceeded under the assumption that ORF2 is the protein responsible for the binding to the BSC element in the screen and in the EMSAs discussed earlier with the A2B extract. Therefore, with help from Kellie Nelson, ORF2 was cloned into the pET-17b expression vector from Novagen (Figure 4) so that fusion protein could be expressed in a bacterial system and subsequently assayed for binding. ORF2 was amplified by PCR using primers TN068 and TN069 to the 3' and 5' ends of ORF2 (see Mouse oligo binder) from the H1 clone in pBSII (pBS/H1) and ligated into pT7Blue and sequenced, pT7/ORF2-13 and pT7/ORF2-16 (Notebook Supplement I, pp.43-57). OFR2 was then digested out of pT7Blue and ligated into the pET-17b vector via BamHI and HindIII, pET/ORF2 (Notebook Supplement I, pp.59-77). Ligation of ORF2 into this vector results in an added 14aa to the N-terminus. Induction was carried out with help from Denise Coberley using the method outlined in the pET System Manual (Novagen). Briefly IPTG was added to bacterial cultures and incubated at 37°C for 15', 30', 1h, 2h, 3h and overnight time intervals. The bacteria were then treated with TritonX-100 and soluble and insoluble fractions collected by keeping the supernatant and pellet, respectively, following sonication and centrifugation. A protein gel run on the pellet and

supernatant fractions indicated that most of the protein was in the pellet fraction and therefore insoluble (Notebook Supplement I, pp. 80-81). The size of the induced protein was approximately 27 kDa. This procedure was repeated at 30°C incubation for 2h and 4h in an attempt to retain the protein in the soluble fraction and in fact this technique worked according to a protein gel on the insoluble and soluble fractions (Notebook Supplement I, pp. 83-84). Hence, we tried an EMSA on this protein extract but the observed shifts were not specific as the same mobility shift was seen in the control lanes also (Notebook VII, pg. 92).

Although the bacterially expressed protein was in the soluble fraction it may not have been folded correctly in this system, therefore we decided to try denaturing and renaturing the protein giving it a chance to refold correctly. 25%, 34%, 40% and 60% cuts of saturated $NH_4SO_4$ were used to precipitate the protein (Schleif and Wensink, pp. 62-64). Following centrifugation the supernatant and pellet were collected and the pellet was resuspended in TE and both were dialyzed against TE. Protein gels after this procedure indicated that some of the protein was in the pellet but that most remained in the supernatant (Notebook Supplement I, pg. 87). Using EMSA analyses on this extract we observed a light shifted band in the 60% and 34% cut extract but we also saw the same mobility shift in the control lane of the 34% cut extract (Notebook VII, pg. 104). Next we tried a mixing experiment where this extract was mixed with mouse embryonic extract and no difference was observed between embryonic extract alone and when mixed with the bacterial fusion protein extract (Notebook VII, pg. 106). Because we had isolated these clones by binding on nitrocellulose and saw no specific binding in the EMSAs with this extract, Rosalyn Juergens and I ran a southwestern analyses to test if ORF2 could bind on nitrocellulose. Both the soluble and insoluble fractions of the 30°C bacterially expressed extract were used and bands of about 27 kDa bound the BSC element in both the soluble and insoluble fractions compared to no binding in the control lanes (Figure 5A).

We then decided to try another bacterial expression system, pGEX (Pharmacia) which would result in the protein being fused to Glutathione Transferase (GST). Dr. C. Tuggle cloned ORF2 and the A4 clone into this vector, made extract and run protein gels (CKT Notebook). Results of the protein gel indicate that most of the induced fusion protein was in the pellet fraction for both clones and that the ORF2 protein was about 27 kDa. Because the protein was again insoluble Dr. C. Tuggle used several different methods to denature and renature the protein to try to solubilize the protein; 6M GuHCl, 8M urea, Tween20 and 0.1% SDS. The protein was then bound to a GST column, washed and eluted. These extracts were then run in an EMSA on which we saw no specific binding compared to a control (CKT Notebook). Next I ran a southwestern where binding was observed in the ORF2 extract, but

not A4, at about 27 kDa (Figure 5B). The southwestern was repeated a second time with the same result (Notebook VII, pg. 152).

A third bacterial expression system designed for retaining soluble, active protein was tried. Dr. C. Tuggle cloned ORF2 into the pET Trx (pET-32c) vector from Novagen. Southwesterns with fusion protein extract made from this expression vector showed binding with ORF2 extract but not with a control vector extract (Figure 5C). The band was approximately 27 kDa. Again, no specific binding was observed in EMSA analyses with ORF2 extract from this system. Lastly, in vitro transcription/translation (Promega) was performed by Dr. C. Tuggle (CKT Notebook) and ORF2 extract from this system did not result in any specific binding in either EMSA (CKT Notebook) or southwestern analyses (Notebook VIII, pp.36 and 40. In conclusion, several attempts to correct the insolubility problem by using different expression systems were unsuccessful. However, southwesterns were positive for binding with all three systems but apparently only to unfused protein. It is therefore, likely that ORF2 cannot bind with extra amino acids fused to the N-terminus. Other evidence for this was seen with the library clones, as none of them were found to actually be fused to β-gal.

Although we failed to see specific binding in EMSAs on all of the extracts except the original fusion protein extract induced directly from the phage incubating on plates, the binding observed in the southwestern with most of these extracts seems to be specific compared to extract from controls. Although ORF2 can bind the BSC element specifically only on "paper" (on the filters during the library screen and in southwestern analyses), we believe ORF2 should be further characterized. ORF2 has no significant homology to anything in the databases (Notebook VII), so at the very least we have identified a new protein. Some protein analysis has been completed on ORF2. Dr. C. Tuggle identified a strong PKC site and one strong and one weak PKA site (Notebook VII). And a protein analysis done by Kellie Nelson using nnpredict and predict protein software predicts a few helical regions but nothing indicative of known DNA-binding motifs (Notebook VII). Therefore, if ORF2 is specifically binding the BSC element it is either due to a motif not predicted by these software programs or due to some unknown binding motif.

We have drawn the following conclusions based on our work thus far:

- induced protein from the A2B clone can specifically bind the BSC element in EMSA analyses
- upon sequencing of the clones, ORF2 is the only complete reading frame in common between the clones

- bacterially expressed ORF2 can specifically bind the BSC element in southwestern analyses
- ORF2 binds only after cleavage from the fusion product and therefore extra amino acids on the N-terminus may interfere with binding and/or folding of this protein
- ORF2 has no homology to anything in the sequence databanks as of February, 1997
- we do not know if this protein binds to the same sites on the BSC element footprinted by embryonic extracts (see second paper)

Future work on ORF2 includes:

- Northern analyses and *in situ* hybridizations to observe RNA expression pattern of ORF2
- antibody production against ORF2 for use in EMSA supershifting experiments to establish whether this protein is part of the complex binding the identified sites on the BSC element
- westerns with ORF2 antibody to observe protein expression pattern of ORF2
- screening of a genomic library with ORF2 to identify the full length gene
- knockout of ORF2 in transgenic mice to observe phenotype and possible altered expression pattern of *Hoxa-5* or other genes (other genes could be based on the expression pattern of ORF2)

**REFERENCES**

Cowell, I. G. and Hurst, H. C. (1993). Cloning transcription factors from a cDNA expression library. In "Transcription Factors: A Practical Approach" (D.S. Latchman, Ed.), pp. 120-122. Oxford University Press, New York.

Huang, S.-H., Tomich, J.M., Wu, H., Jong, A. and Holcenberg, J. (1989). Human Deoxycytidine Kinase. J. Biol. Chem. 264:14762-14768.

Huang, S.-H. and Jong A. (1994). Efficient induction and preparation of fusion proteins from recombinant phage λgt11 clones (Technical Tips). TIG 10:183.

Schleif, R.F. and Wensink, P.C. (1981). Working with proteins. In "Practical Methods in Molecular Biology" (P. Manor, Ed.), pp. 62-64. Springer-Verlag, New York, Inc.

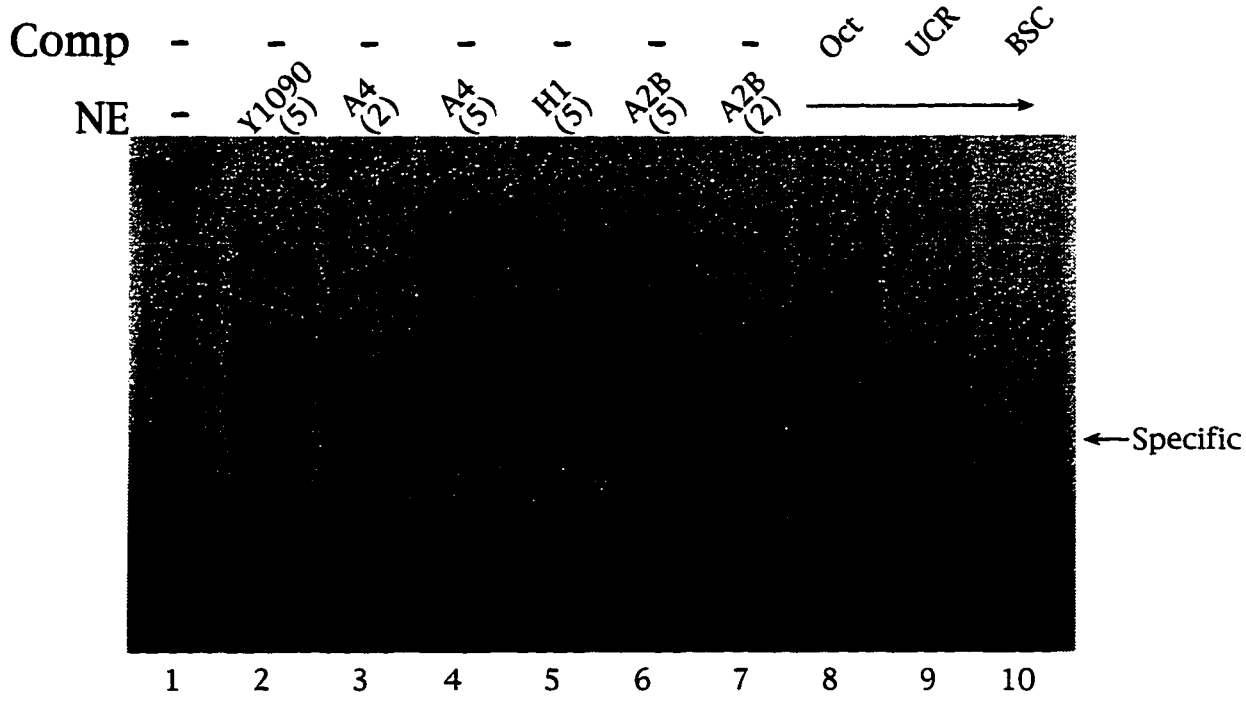



Figure 1. EMSA analysis of binding to the BSC element in extract prepared from cDNA-containing phage. cDNA sequences were isolated from an expression library screen with the BSC element as probe and protein was induced directly from the growing phage plaques. Lane 1 contains no extract and lane 2 contains extract from uninfected Y1090 cells. The following lanes contain extract from phage containing the various clones: lane 3 and 4- clone A4 induced for 2 or 4 hours, A4 (2) and A4 (5); lane 5- clone H1 induced for 5 hours, H1 (5); lane 6- clone A2B induced for 5 hours, A2B (5), and lanes 7-10- clone A2B induced for 2 hours. Lanes 8-10 also contain the following competitors: lane 8- oct oligo as nonspecific competitor (Oct); lane 9- upper cervical repressor element as nonspecific competitor (UCR); and lane 10- the brachial spinal cord element as specific competitor (BSC).

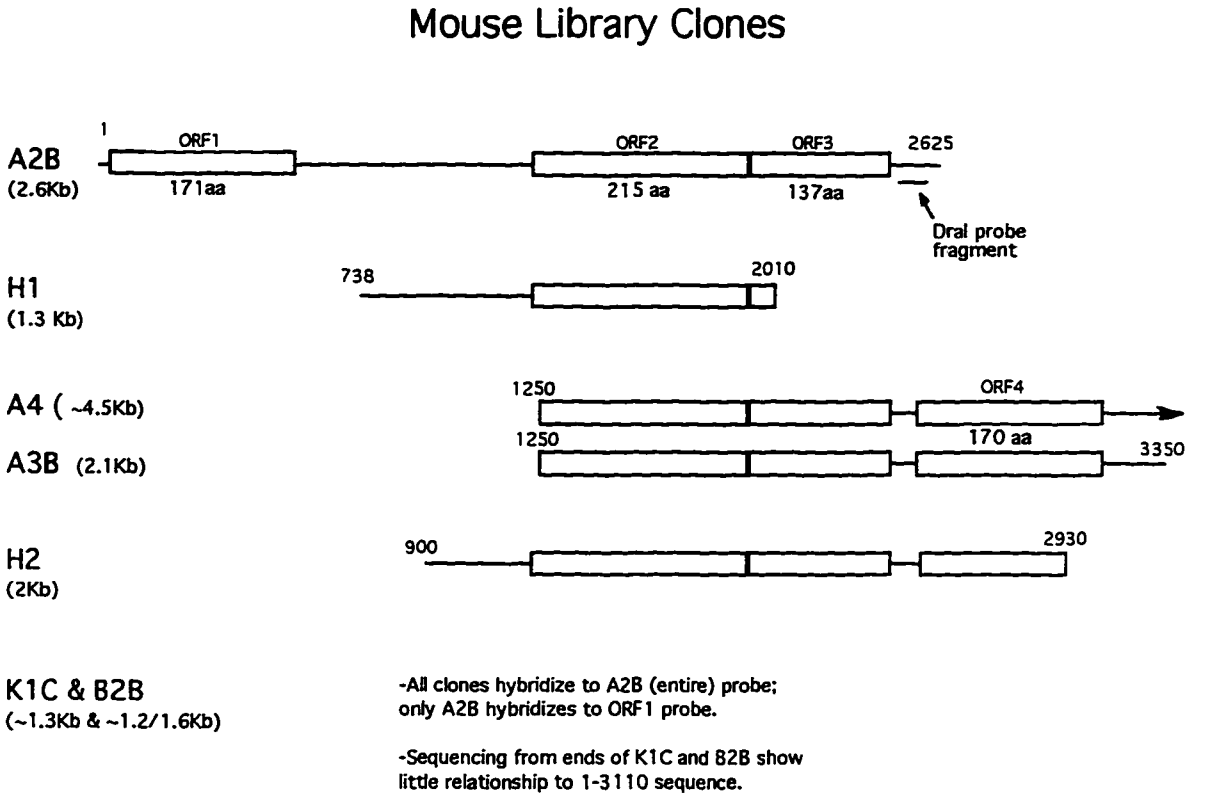



Figure 2. Graphic depiction of the isolated library clones showing the sizes and sequences in common between the clones. Also shown are the locations of identified open reading frames (ORF). Complete sequence has been obtained from the beginning of clone A2B (1bp) through the end of A3B (3350bp).

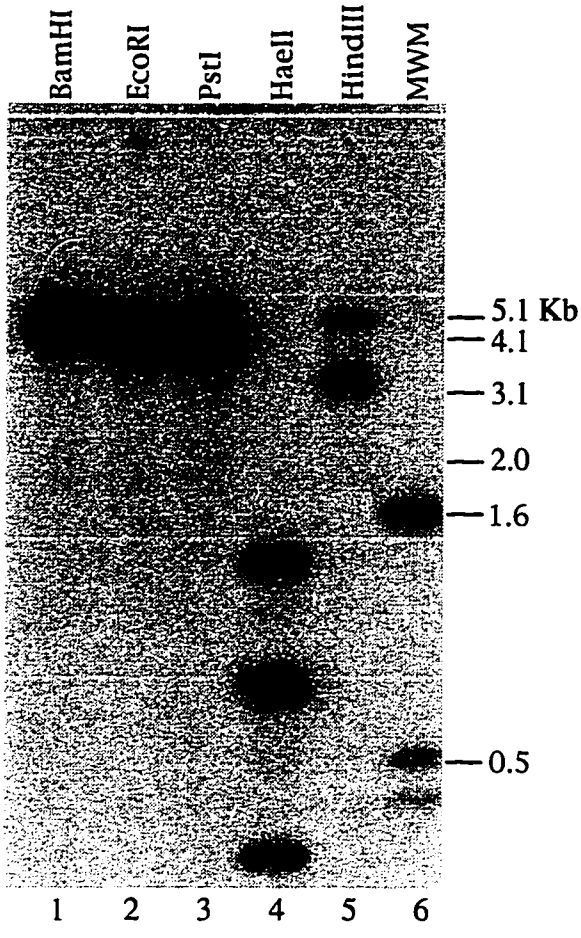



Figure 3. Southern blot containing genomic mouse DNA digested with the following enzymes; lane 1- BamHI, lane 2- EcoRI, lane 3- PstI, lane 4- HaeII, and lane 5- HindIII and hybridized to the entire H1 clone. MWM lane containing the 1 Kb ladder (Boehringer Mannheim) is located in lane 6 and sizes are indicated to the right of this lane.

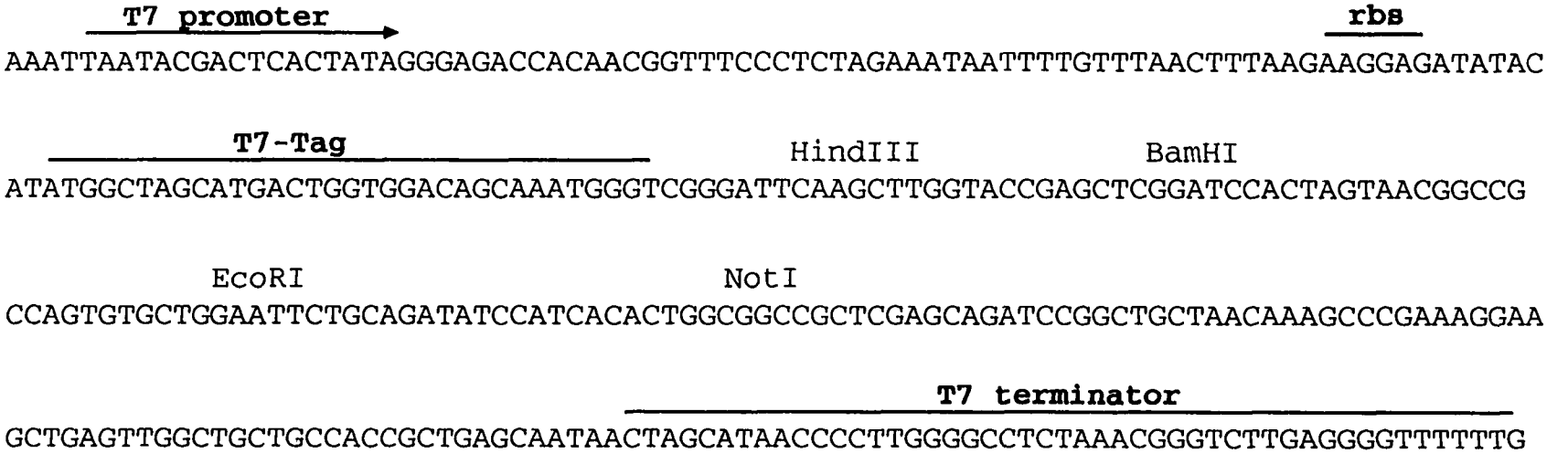





Figure 4. pET-17b cloning/expression region. ORF2 was inserted via BamHI/HindIII in frame with the first methionine located atp the beginning of the T7-Tag region for use in producing a fusion protein in bacteria. rbs- ribosome binding site.

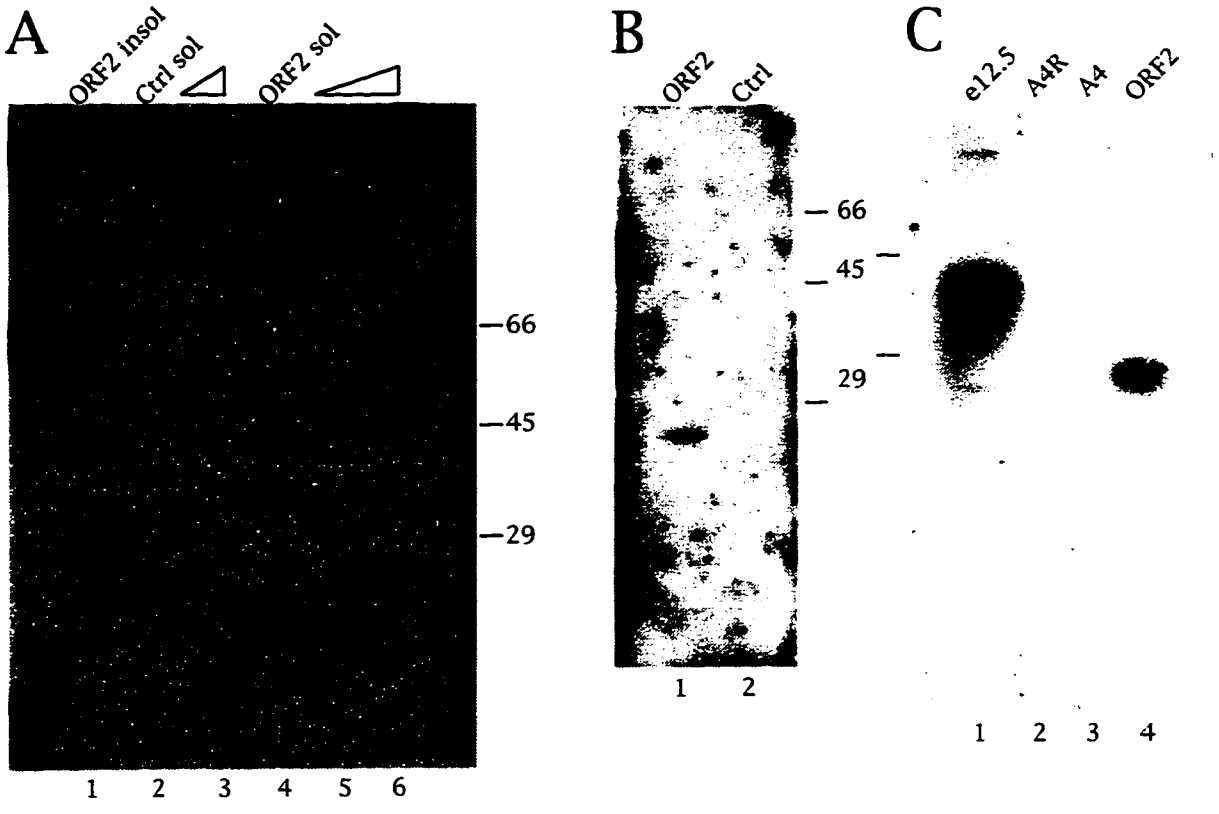



Figure 5. Protein blots containing fusion protein extract from pET-17b (A), pGEX (B), or pET-32c (pTrx) (C) bacterial expression vectors. Blots were hybridized with the BSC element. (A) Lane 1- extract from the insoluble fraction of pET/ORF2 induced at 37°C (ORF2 insol), lane 2 & 3- extract from pET-17b empty vector soluble fraction induced at 30°C (Ctrl sol), lane 4-6- extract from the soluble fraction of pET/ORF2 induced at 30°C (ORF2 sol). (B) Lane 1- extract from pGEX/ORF2 (ORF2) and lane 2- extract from pGEX empty vector (Ctrl). (C) Lane 1- nuclear extract from day 12.5 mouse embryos (e12.5), lane 2- extract from pTrx/A4 in the reverse orientation (A4R), lane 3- extract from pTrx/A4 (A4), and lane 4- extract from pTrx/ORF2 (ORF2).

# APPENDIX II
# PROTOCOLS

## Southwestern

(Reference: "Transcription Factors: A Practical Approach", pp. 110-112, 1993)

1) Protein Gel- Run samples on gel:

a) 25-50 µg of Nuclear Extract in 1X sample buffer, boil samples @95°, 5'

b) cool samples to room temp, spin briefly and load on gel in cold room (4° C)

c) run gel in cold 1X Tank Buffer

d) ~6-7 hours at 35mA, constant current on 12% running gel (or 200 volts, constant voltage, at least 6-7 hours)

2) Transfer to Nitrocellulose

a) Cut 0.45µm nitrocellulose and 2 pieces of Whatman No.1 to size of running gel

b) Equilibrate gel, nitrocellulose, and Whatman paper in cold Transfer Buffer, 30'

c) Assemble Trans-Blot apparatus according to directions (Gray panel-fiber pad-filter paper-gel-nitrocellulose-filter paper-fiber pad-clear panel) Gray panel faces black side of apparatus. Hint: When assembling, mark top and bottoms of lanes with nicks and place notches at top to designate left and right. Also mark with nitrocellulose marker protein/non-protein side

d) Fill apparatus with cold Transfer Buffer, place stir bar in bottom and place on stir plate

e) Place black coil with cold circulating water running through it in apparatus (the circulating bath will have been previously cooled to 0-4°).

f) Transfer at 250 mA, constant current, 2 hours with constant stirring.

3) Blocking of Membranes

a) Cut off MWM lanes and stain in Amido Black (5-10') and destain

b) Equilibrate rest of nitrocellulose in TNE-50 briefly

c) Block nitrocellulose in SW Block (250 ml) O/N, 4°, on shaking platform

4) Hybridization

a) Rinse nitrocellulose in TNE-50 briefly (cut membrane into strips if necessary, following the nicks made earlier as guidelines)

b) Place membranes in plastic ziplock baggies with 10-20 ml TNE-50 (depending on size of membrane) + polyanion (an appropriate amount- use ~8μg per blot of poly d(A/C)(G/T) when probing with the 38-6 BSC element) and incubate at RT, 5'

c) Add probe to respective baggies (~2-5 million counts) and incubate at RT, 1 hour, on shaking platform (flip baggie at 30')

d) Dump off hyb soln and wash each membrane 3-4x in 20 ml TNE-50, 10', RT, checking membranes with Geiger counter between washes

e) wrap each membrane with saran wrap and expose to film

Solutions:

1) 4X Tank Buffer-

| | |
|---|---|
| 24g | Tris base |
| 115.2g | Glycine |
| 8g | SDS |

Bring vol to 2L with $ddH_2O$
pH should be 8.3
Use 500ml in 2L for running Protein Gel

2) Transfer Buffer-

| | |
|---|---|
| 12.2g | Tris base |
| 57.6g | Glycine |

Bring vol to 2L with $ddH_2O$

| | |
|---|---|
| 800ml | Methanol |

Bring vol to 4L with $ddH_2O$

3) TNE-50-

| | |
|---|---|
| 10ml | 1M Tris, pH 7.5 |
| 10ml | 5M NaCl |
| 2ml | 0.5M EDTA |
| 1ml | 1M DTT |

Bring vol to 1L with $ddH_2O$

4) SW Block-

| | |
|---|---|
| 6.25g | Dried Milk |
| 50ml | 125mM HEPES, pH 8.0 |
| 250μl | 1M DTT |
| 25ml | 100% Glycerol |
| 2.5ml | 5M NaCl |
| 0.5ml | 0.5M EDTA |

Bring vol to 250ml with $ddH_2O$

5) 5X Sample Buffer-

| | |
|---|---|
| 0.154g | DTT |
| 0.2g | SDS |
| 0.8ml | 1M Tris, pH 6.8 |

Dissolve the above

0.1ml Glycerol
Pinch of Bromophenol Blue

**Nuclear Extract Preparation**

(Reference: Roy et al. BioTechniques 11:770-777, 1990)

1) Isolate 0.5-1g of tissue or $5x10^7$ - $1x10^8$ cells from cell culture

2) Homogenize 10 strokes in a 7ml glass Dounce homogenizer in 4ml NE1 Buffer

3) Filter through 2 layers of sterile cheesecloth

4) Add NP-40 to final concentration of 0.5%

5) Homogenize another 5 strokes

6) Pellet nuclei and cell debris in Sorvall 8', 1000xg, 4°

7) Wash pellet with 4-5ml NE1 Buffer, centrifuge as in #6

8) Discard supernatant and lyse nuclei in 1PCV (Packed Cell Vol) of NE2 Buffer, 5', 4°

9) Add 1/10 PCV of 4M KCl

10) Homogenize 20 strokes

11) Transfer to 1.5ml microcentrifuge tube, centrifuge 5', 12000xg, 4°

12) Transfer supernatant to "special" thick-walled 1.5ml tube and ultracentrifuge at 100000xg, 90', 4° (70k with 300.1 rotor, tabletop ultracentrifuge in Eric Henderson's lab)

13) Dialyze supernatant against DNaseI Buffer, 1hr, 4°

14) Spin down any denatured proteins after dialysis, aliquot supernatant, quick freeze. and store at -70°

Solutions:

| NE1 Buffer | Stocks | NE2 Buffer |
|---|---|---|
| 28.4ml | 30% sucrose | 2.8ml |
| 1.5ml | 1M Tris, pH7.9 | 150µl |
| 2.8ml | 5M NaCl | 280µl |
| 400µl | 0.5M EDTA | 40µl |
| 156µl | 0.32M EGTA | 15.6µl |
| 832µl | 3M KCl | 1.2ml |
| 200µl | 1M $MgCl_2$ | 20µl |
| 150µl | 0.1M Spermine | 15µl |
| 500µl | 0.1M Spermidine | 50µl |
| 100µl | 1M DTT | 10µl |
| 400µl | 0.1M PMSF | 40µl |
| 250µl | 1mg/ml Pepstatin | 25µl |
| 50µl | 5mg/ml Leupeptin | 5µl |
| 25µl | 10mg/ml Aprotinin | 2.5µl |
| to 100ml | $ddH_2O$ | to 10ml |

| | | |
|---|---|---|
| DNaseI Buffer- | 33.4ml | 3M KCl |
| | 8ml | 1M $MgCl_2$ |
| | 40ml | 1M $K_3PO_4$, pH 7.4 |
| | 2ml | 1M β-mercaptoethanol |
| | 400ml | 100% Glycerol |
| | Bring vol. to 2L with $ddH_2O$ | |

**DNaseI Footprinting**

(Reference: "Transcription Factors: A Practical Approach", pp. 32-40, 1993)

1) Pour a 7%, 1XTBE, 8M Urea, 19:1 Acryl:Bisacryl sequencing gel

2) Dilute RQ1 DNase (1U/μl)- 5μl in 95μl of cold Tris, pH 8.0 for a final conc. of 0.05U/μl

3) For the first run the amount of DNase needed for optimum digestion should be determined using only the labeled probe, for example:

| | |
|---|---|
| 25μl | 2X Binding Buffer (BB) |
| 20μl | $dH_2O$ (at 4°C) |
| 5μl | probe |

Several tubes should be made up and only 2-3 tubes should be processed at one time.

Once the amount of DNase needed has been determined empirically, binding reactions with nuclear extract should be set up as in an EMSA, beginning with no extract and then with increasing amounts of extract. For example:

| Rxns | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 5X BB | 10μl | 10 | 10 | 10 | 10 |
| $dH_2O$ | 35 | 26.7 | 18.4 | 10.1 | 1.8 |
| NE | 0 | 5 | 10 | 15 | 20 |
| polyanion | 0 | 3.3 | 6.6 | 9.9 | 13.2 |
| probe | 5 | 5 | 5 | 5 | 5 |

NE= Fractionated whole e12.5 embryos (7.5μl was able to completely shift 0.5ng of probe in EMSA)

polyanion= d(A/C)d(G/T), 400ng/μl

probe= either T/H or E/S fragments from pJ1, 0.1ng/μl, 4600 cnts/μl

4) Mix and incubate the reactions RT, 20' (as in EMSA). Heat stop solution to 37°C- mix well.

5) Add 50μl RT $Ca^{2+}/Mg^{2+}$ solution- incubate 1', RT

6) Immediately add 10μl RQ1 (previously diluted), incubate 5', RT (empirically determined) Discard unused diluted RQ1 DNase

7) Add 90μl stop soln (previously heated to 37°C), vortex

8) Add 200μl Phenol:Chloroform (1:1), spin 5'

9) To top layer add 500µl 100% EtOH, incubate 20' on ice to ppt, spin 10'

10) Wash pellet with 1ml 70% EtOH, vortex, spin 5'

11) Resuspend pellet in 4µl loading solution (vortex and quick spin)

12) Heat rxn 2', 95 °C- chill on ice, quick spin

13) Load 3µl on gel (prerun at 60 Watts, constant power, ~45')

14) Run gel at 60 Watts, constant power (~1.5 hrs)

15) Fix and dry gel, expose to film

Solutions:

1) Sequencing Gel-

Stock:

| | |
|---|---|
| 145ml | 95.2g Urea in 145ml $dH_2O$ (heat to dissolve) |
| 35ml | 40% 19:1 Acryl:Bisacryl |
| 20ml | 10X TBE |

Filter and store out of light

For pouring gel:

| | |
|---|---|
| 90ml | Of above mixture |
| 0.9ml | 10% APS |
| 20µl | TEMED Stop Solution- |

| | |
|---|---|
| 200mM | NaCl |
| 30 mM | EDTA |
| 1% | SDS |
| 100µg/ml | yeast tRNA |

2) Loading Solution-

| | |
|---|---|
| 0.1M | NaOH:Formamide (1:2, v/v) |
| 0.1% | xylene cyanol |
| 0.1% | bromophenol blue |

3) $Ca^{2+}/Mg^{2+}$ Solution-

| | |
|---|---|
| 5mM | $CaCl_2$ |
| 10mM | $MgCl_2$ |

4) 5X Binding Buffer (BB)-

| | |
|---|---|
| 400µl | 125mM HEPES, pH 7.9 (25mM final) |
| 1ml | 100% Glycerol (50% final) |
| 83.3µl | 3M KCl (125mM final) |
| 1µl | 0.5M EDTA (0.25mM final) |
| 12.5µl | 100mM PMSF (0.625 mM final) |

Bring vol to 2ml with $dH_2O$

## Expression Library Screen

I. Titering Library

A. Dilutions:

1. 5µl library in 500µl SM Buffer ($10^{-2}$)

2. 10µl of #1 in 1ml SM Buffer ($10^{-4}$)
3. 100µl of #2 in 1ml SM Buffer ($10^{-5}$)
4. 100µl in #3 in 1ml SM Buffer ($10^{-6}$)

B. Incubate 100µl of #2-4 and 200µl Y1090 cells (O/N culture grown with Amp) in separate falcon tubes 20min at 37°C

C. Add 3ml 53°C top agarose, mix and pour on prewarmed LB plates (100mm, small plates)

D. Incubate O/N at 37°C and count plaques the following morning to determine the titer [number of plaque forming units (pfu) in library]

II. Plating of Library

Plate $5x10^4$ pfu per large plate for a total of at least 500,000 clones (10-12 plates):

Incubate 200µl of diluted library and 500µl Y1090 cells (O/N culture grown with Amp) in separate falcon tubes 20min at 37°C

Add 8ml 53°C top agarose, mix and pour on prewarmed LB plates (150mm, large plates) and incubate at 42°C ~3h (until tiny plaques are visible)

III. Replica Lifts

A. Soak 132mm nitrocellulose filters (0.45µm- Nitrobind from Micron Separations Inc.) in 10mM IPTG (100ml for 24 filters) 30min and let dry on Whatman paper

B. Number the filters and plates. Make 3 nonsymmetrical dots on the bottom of the plates

C. Lay filters on the plates and make pinpricks through the nitrocellulose and agar directly over where the dots are on the bottom of the plate (these will be used to line up the films later in order to select positives)

D. Incubate filters on the plates 1h, 37°C

E. Remove filters with forceps to a box containing TNE-50 buffer

F. Place second set of filters on plates (marking them the same as the first set of filters) and incubate 3h, 37°C

G. Wash filters 5-10min in TNE-50 Buffer

H. Block O/N in 250ml SW Block, 4°C, with gentle shaking

IV. Probing of Filters

A. Transfer the first set of filters into TNE-50 in plastic box for probing (6 filters per box) and wash briefly, dump buffer

B. Add 50ml TNE-50 + polyanion to box and incubate 10min, RT

C. Add probe previously labeled at $5x10^5$/ml (~$2.5x10^7$ per 50ml) and incubate 1h, RT, with gentle shaking. At 30min change the order of the filters.

D. Remove filters with forceps into new box and wash with TNE-50. All 12 filters can be washed in 250ml TNE-50, 10min, 3-4x, RT, with shaking- checking counts after each wash.

E. Probe the second set of filters (6 per box) in probe solution, 1.5h, RT, with shaking, and wash as above.

F. Place filters between saran wrap and expose film (try to arrange so duplicates are together and oriented the same way, use 2 Stratagene stickers for orientation of film on filters).

V. Selection of Positives

A. Line up film on the filters (using the Stratagene stickers) to mark on the film where the pinpricks are.

B. Cut the film apart into individual filters and line up film on the plates and circle on the bottom of the plates the positive plaques, seen as duplicates on both sets of filters.

C. Pick the plaques using the large bore end of Pasteur pipettes. Place the agar plugs in 500µl SM Buffer O/N, 4°C to resuspend the phage. (Place one drop, ~5µl, chloroform in each tube and mix to ppt bacteria)

D. Spin down plugs and bacteria and remove supernatant to new tube, store resuspended phage at 4°C.

VI. Secondary (and Tertiary) Screen

A. Titer resuspended positive phage from primary (secondary) screen as in I.

B. Plate ~600 pfu per large plate and screen as done previously.

C. Tertiary screen can be done on small plates, plate ~50-200pfu/plate (use 82mm filters).

D. Can resuspend positives from 2° and 3° screens in $dH_2O$ and use PCR to look at sizes using 5' and 3' gt11 primers

Solutions:

1) TNE-50-

| | |
|---|---|
| 10ml | 1M Tris, pH 7.5 |
| 10ml | 5M NaCl |
| 2ml | 0.5M EDTA |
| 1ml | 1M DTT |

Bring vol to 1L with $ddH_2O$

2) SW Block-

| | |
|---|---|
| 6.25g | Dried Milk |
| 50ml | 125mM HEPES, pH 8.0 |
| 250µl | 1M DTT |
| 25ml | 100% Glycerol |
| 2.5ml | 5M NaCl |
| 0.5ml | 0.5M EDTA |

Bring vol to 250ml with $ddH_2O$

3) SM Buffer

| | |
|---|---|
| 5.8g | NaCl |
| 2g | $MgSO_4$-$7H_2O$ |
| 50ml | 1M Tris-HCl, pH 7.5 |
| 5ml | 2% gelatin |

Bring vol to 1L with $dH_2O$
sterilize by autoclaving. Store in 50ml lots

4) Top Agarose
7g Agarose added to 1L LB before autoclaving to sterilize

Other Needed Supplies:
Nitrocellulose filters: 132mm and 82mm
Whatman
IPTG
LB agar plates: 150mm and 100mm
X-ray Film

## ACKNOWLEDGMENTS

I would like to thank Howard Hughes students Rosalyn Juergens, Dianna Wirzbach, and especially Kellie Nelson and graduate students Michelle Aulik and Diana Walker for their hard work and assistance while rotating through the lab. I would also like to thank Claire Schmitz, Tun-Ping Yu, Kari Beth Krieger, and Melissa Allen for anything and everything they were so willing to help me with and especially for their friendship. And thank you to my advisor, Dr. Chris Tuggle, for his support and guidance during my studies.